

# The Capacity Region of Broadcast Channels with Intersymbol Interference and Colored Gaussian Noise

Andrea J. Goldsmith, *Senior Member, IEEE*, and Michelle Effros, *Member, IEEE*

**Abstract**—We derive the capacity region for a broadcast channel with intersymbol interference (ISI) and colored Gaussian noise under an input power constraint. The region is obtained by first defining a similar channel model, the circular broadcast channel, which can be decomposed into a set of parallel degraded broadcast channels. The capacity region for parallel degraded broadcast channels is known. We then show that the capacity region of the original broadcast channel equals that of the circular broadcast channel in the limit of infinite block length, and we obtain an explicit formula for the resulting capacity region. The coding strategy used to achieve each point on the convex hull of the capacity region uses superposition coding on some or all of the parallel channels and dedicated transmission on the others. The optimal power allocation for any point in the capacity region is obtained via a multilevel water-filling. We derive this optimal power allocation and the resulting capacity region for several broadcast channel models.

**Index Terms**—Broadcast channels, capacity region, colored Gaussian noise, intersymbol interference (ISI).

## I. INTRODUCTION

TWO of the most celebrated results in Shannon theory are Gallager's capacity and corresponding water-filling formula for single-user Gaussian channels with intersymbol interference (ISI) and colored noise [1] and Cover's superposition coding technique for degraded broadcast channels [2]. In this paper, we show that the capacity region of a Gaussian broadcast channel with ISI and colored noise is achieved using a strategy that combines superposition coding with an optimal power allocation achieved by water-filling.

The channel model we consider consists of one transmitter sending information to two receivers, each with a different finite-channel impulse response and colored Gaussian noise of finite memory. Given an average power constraint, the transmitter may send independent information to each receiver or a combination of independent information for each receiver and common information intended for both receivers. We obtain the capacity region in both cases, and for independent information our results can be easily extended to any finite number of users. The capacity

regions define the maximum set of rates that can be transmitted to multiple users with arbitrarily small error probability. We also obtain numerical results for the capacity region and optimal power allocation on several example broadcast channels with independent information.

We derive the capacity region of the broadcast channel with ISI and colored noise using the same methodology that is used to obtain the capacity of the single-user and synchronous multiple-access channels (MACs) with ISI [3]–[5]. Specifically, we first show that our broadcast channel model has the same capacity region as a circular broadcast channel model that can be decomposed using the discrete Fourier transform (DFT), and we then obtain the capacity of this circular model based on the DFT decomposition. Our approach differs from those of [3]–[5] in three significant ways. First, we use different techniques to show that the circular channel model has the same capacity region as the original channel. Specifically, since there is no known formula for the capacity region of a general broadcast channel with memory, we show that the circular and original broadcast channels have the same capacity region using only the definitions of these regions. In fact, we prove a more general result: Any synchronous multiterminal channel [6, Sec. 14.10] and its circular approximation have the same capacity region. Applying this general result to the single-user and synchronous multiple-access channels provides alternative proofs for [3, Corollary 1] and [4, Theorem 1]. We also explicitly treat the effect of ISI and noise correlation between codeword transmissions, which has not been done previously for any multiuser channel with memory. These effects impact any system with back-to-back codeword transmission, as is typical in practice. For such systems, ISI causes interference between subsequent codewords, and therefore may impact the optimal codeword design. Also, noise correlation causes subsequently received codewords to be correlated, and therefore may impact the optimal decoder design. Gallager has shown that ISI and noise correlation between codewords does not impact channel capacity for single-user channels with finite memory, since these channels are indecomposable [1, Theorem 4.6.4]. There is no such parallel theory for broadcast or multiple-access channels. Therefore, we modify the standard definitions of error probability and achievable rate for multiterminal channels with finite memory to account for ISI and noise correlation between codeword transmissions.<sup>1</sup> We then show that the capacity region for finite-memory multiterminal channels under these definitions is independent of previous codeword transmissions and noise samples and equals that of the multiterminal circular channel. This general result is applied to the

Manuscript received September 16, 1998; revised December 1, 1999. The work of A. J. Goldsmith was supported by an ONR Young Investigator Award N00014-99-1-0698 and by an NSF CAREER Award NCR-9501452. The work of M. Effros was supported by an NSF CAREER Award MIP-9501977, a grant from the Powell Foundation, and donations from Intel.

A. J. Goldsmith is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94025-9515 USA (e-mail: andrea@ee.stanford.edu).

M. Effros is with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125 USA (e-mail: effros@caltech.edu).

Communicated by M. L. Honig, Associate Editor for Communications.

Publisher Item Identifier S 0018-9448(01)00461-8.

<sup>1</sup>Our modified capacity definition is basically an extension of Wolfowitz's single-user capacity definition [7, Sec. 6] that takes the channel's past transmissions and noise into account.

broadcast channel to show it has the same capacity region as the circular broadcast channel. We then derive the circular broadcast channel capacity region, and this derivation differs from those in [3]–[5] since it is based on the capacity region of parallel degraded broadcast channels [8]–[10].

The derivation of the circular broadcast channel capacity region is obtained through a DFT decomposition to give a set of parallel independent degraded broadcast channels in the frequency domain, for which the capacity is known. The achievability of the multiuser capacity region for a parallel set of degraded broadcast channels with independent information, along with the corresponding coding strategy and optimal power allocation, was obtained by Hughes-Hartogs in 1975 [8]. The converse proof for this capacity region in the two-user case was obtained by Poltyrev in [9], and the two-user capacity region for the parallel degraded broadcast channel with common information was derived by El Gamal in [10]. We make use of these previous results to obtain the capacity region, optimal coding strategy, and optimal power allocation for our broadcast channel model.

The coding strategy that achieves capacity on the broadcast channel with ISI and colored noise uses superposition coding at some or all frequencies and dedicated transmission to a single user at the others. For two users with independent information, the optimal power allocation between users on each channel is obtained graphically via a multilevel water-filling, which assigns different water levels to the different channels depending on the relative priorities of the users. This multilevel water-filling is reminiscent of Gallager's water-filling strategy for a single-user channel with ISI and colored noise [1]. We derive this optimal water-filling strategy and the corresponding capacity region for several example channels. In practice, this optimal power allocation and coding strategy could be implemented using multicarrier modulation, which also uses a DFT to decompose an ISI channel into parallel channels in the frequency domain.

The remainder of this paper is organized as follows. In Section II, we define the channel of interest, its circular approximation, and a memoryless block channel derived from the original channel. In Section III, we define the capacity region for these channels and establish our notation. In Section IV, we show that all three of these channel models have the same capacity region. Exact formulas for the capacity region based on a DFT decomposition of the circular channel, given by (32) and (33), are obtained in Section V. The capacity region and optimal power allocation strategies for several example channel models are given in Section VI, followed by our conclusions.

## II. CHANNEL MODELS

We consider the discrete-time broadcast channel model shown in Fig. 1. The model consists of one transmitter sending information to multiple receivers. For simplicity, we only consider the two-user broadcast channel, but our results can be extended to any finite number of users.<sup>2</sup> Throughout the paper,

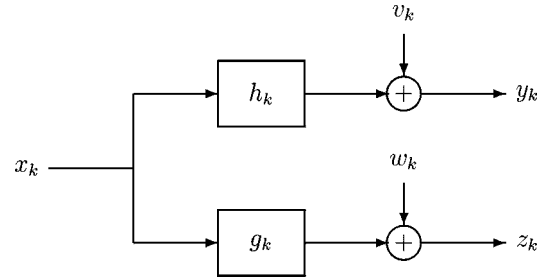


Fig. 1. Two-user discrete Gaussian broadcast channel with ISI.

we denote subsequence  $(s_a, \dots, s_b)$  as  $\{s_k\}_{k=a}^b$ , sequence  $(\dots, s_1, s_2, \dots)$  as  $\{s_k\}$ , and vector  $(s_1, \dots, s_n)$  as  $s^n$  for any  $s$ .

In the model of Fig. 1, the real input sequence  $\{x_k\}$  is sent to each user over different real time-invariant channels with additive Gaussian noise. The stationary noise processes, denoted by  $\{v_k\}$  and  $\{w_k\}$ , have mean zero and autocorrelation functions  $R_v[i]$  and  $R_w[i]$ , respectively. We assume that these autocorrelation functions have finite support so that for some finite  $i_{\max}$ ,  $R_v[i] = R_w[i] = 0$  for  $|i| \geq i_{\max}$ . We let  $\{h_i\}_{i=0}^m$  and  $\{g_i\}_{i=0}^m$  denote the real finite impulse response of the first and second channels, respectively, with  $m$  the common finite impulse response duration of both channels. We consider cases where  $m \geq i_{\max}$  only, and  $m$  is defined as the broadcast channel memory.<sup>3</sup> The real input sequence  $\{x_k\}$  produces the real output sequences  $\{y_k\}$  at the first receiver and  $\{z_k\}$  at the second receiver, with

$$y_k = \sum_{i=0}^m h_i x_{k-i} + v_k = h_k * x_k + v_k, \quad (1)$$

and

$$z_k = \sum_{i=0}^m g_i x_{k-i} + w_k = g_k * x_k + w_k \quad (2)$$

where  $*$  denotes linear convolution. The channels have ISI since the channel output at time  $k$  depends on the input symbol  $x_k$  as well as previous input symbols  $x_i$ ,  $i < k$ . In particular,  $y_k$  and  $z_k$  for  $k < m$  depend on inputs  $x_i$  for  $i < 1$ , i.e., on previous codeword transmissions. In addition, for  $k < m$ , the noise samples  $v_k$  ( $w_k$ ) are correlated with the noise samples  $v_i$  ( $w_i$ ) for  $i < 1$ , so  $y_k$  ( $z_k$ ) is also correlated with these noise samples.

The transfer function of the first receiver's channel is

$$H(\omega) = \sum_{i=0}^m h_i e^{-ji\omega} \quad (3)$$

and the transfer function of the second receiver's channel is

$$G(\omega) = \sum_{i=0}^m g_i e^{-ji\omega}. \quad (4)$$

<sup>2</sup>This extension requires the capacity region for a set of parallel degraded broadcast channels with multiple users. The achievability of this region with independent information was derived in [8, Ch. 4], and the converse can be obtained by extending the two-user converse in [9]. The capacity region for parallel broadcast channels with common information and more than two users is not known.

<sup>3</sup>If for  $m \geq i_{\max}$  one of the channels has an impulse response shorter than  $m$ , then we can pad this response with the appropriate number of zeros to obtain a filter that satisfies the desired constraint. Similarly, if both channels have an impulse response duration less than  $i_{\max}$ , then we set  $m = i_{\max}$  and pad both responses with the appropriate number of zeros.

The spectral noise density of the first receiver's channel is

$$N_v(\omega) = \sum_{i=-(m-1)}^{m-1} R_v[i]e^{-ji\omega} \quad (5)$$

and the spectral noise density of the second receiver's channel is

$$N_w(\omega) = \sum_{i=-(m-1)}^{m-1} R_w[i]e^{-ji\omega}. \quad (6)$$

We assume an average power constraint so that for all  $n$

$$\frac{1}{n} \sum_{k=1}^n E[X_k^2] \leq P \quad (7)$$

where the expectation is taken with respect to the distribution of the randomly chosen codeword  $(X_1, \dots, X_n)$ . For a given  $m$  we call this channel the linear Gaussian broadcast channel (LGBC) with finite memory  $m$  and average power  $P$ . We define the initial state of this channel as  $\mathcal{S}_0 = (\mathcal{X}_0, \mathcal{V}_0, \mathcal{W}_0)$ , where

$$\mathcal{X}_0 = (x_{-m+1}, \dots, x_0)$$

$$\mathcal{V}_0 = (v_{-m+1}, \dots, v_0)$$

and

$$\mathcal{W}_0 = (w_{-m+1}, \dots, w_0).$$

The vector  $\mathcal{X}_0$ , which determines the ISI from the previous codeword transmission to the current codeword transmission, impacts the first  $m$  channel outputs  $\{y_k\}_{k=1}^m$  and  $\{z_k\}_{k=1}^m$  in our model. Similarly, the vectors  $\mathcal{V}_0$  and  $\mathcal{W}_0$  affect the first  $m$  noise samples (and thereby the first  $m$  channel outputs) at receivers 1 and 2, respectively, due to noise correlation.

We now define two additional  $n$ -block memoryless channel models derived from the LGBC. A broadcast channel is  $n$ -block memoryless if for any integer  $K$

$$p(y^{Kn}, z^{Kn} | x^{Kn}) = \prod_{k=1}^K p(y_{(k-1)n+1}^{kn}, z_{(k-1)n+1}^{kn} | x_{(k-1)n+1}^{kn}). \quad (8)$$

By (8) we see that the channel outputs  $(y_i, z_i)$  over any  $n$ -block transmission are independent of channel inputs  $x_i$  and noise samples  $v_i$  and  $w_i$  from previous or subsequent  $n$ -block transmissions. In particular, if for any integer  $k$  we send codewords of length  $kn$  over an  $n$ -block memoryless channel then the outputs  $(y_i, z_i)$ ,  $1 \leq i \leq n$ , are independent of any initial channel state  $\mathcal{S}_0$ . Note, however, that if we send codewords of length  $l$  over an  $n$ -block memoryless channel, where  $l$  is not an integer multiple of  $n$ , then the channel outputs corresponding to one codeword transmission will be affected by the previous codeword transmission, since the  $n$ -block memoryless channel is not  $l$ -block memoryless.

The  $n$ -block circular Gaussian broadcast channel ( $n$ -CGBC), defined for  $n > m$ , is an  $n$ -block memoryless channel obtained by modifying the LGBC with memory  $m$ . Specifically, the  $n$ -CGBC over each  $n$ -block has real input vector  $\{x_k\}_{k=1}^n$

which produces the real output vector  $\{\tilde{y}_k\}_{k=1}^n$  at the first receiver and  $\{\tilde{z}_k\}_{k=1}^n$  at the second receiver with

$$\tilde{y}_k = \sum_{i=0}^m h_i x_{((k-i))_n} + \tilde{v}_k = h_k \otimes x_k + \tilde{v}_k \quad (9)$$

and

$$\tilde{z}_k = \sum_{i=0}^m g_i x_{((k-i))_n} + \tilde{w}_k = g_k \otimes x_k + \tilde{w}_k \quad (10)$$

where  $\otimes$  denotes circular convolution and  $((s))_n$  equals  $s$  modulo  $n$  except when  $s$  is zero or an integer multiple of  $n$ , in which case  $((s))_n = n$ . This modification of the standard modulo  $n$  definition results from our definition that  $n$ -blocks  $\{s_k\}_{k=1}^n$  start at time  $k = 1$  instead of  $k = 0$ . Performing the corresponding circular convolution on each previous and subsequent  $n$ -block yields the remainder of the output sequences  $\{\tilde{y}_k\}$  and  $\{\tilde{z}_k\}$ . The noise processes over each  $n$ -block  $\{\tilde{v}_k\}_{k=1}^n$  and  $\{\tilde{w}_k\}_{k=1}^n$  are defined as stationary Gaussian processes with

$$E[\tilde{v}_k] = E[\tilde{w}_k] = 0$$

$$E[\tilde{v}_k \tilde{v}_l] \triangleq \tilde{R}_v[k-l] = R_v[k-l] + R_v[k-l+n] + R_v[k-l-n]$$

and

$$E[\tilde{w}_k \tilde{w}_l] \triangleq \tilde{R}_w[k-l] = R_w[k-l] + R_w[k-l+n] + R_w[k-l-n].$$

Thus, the noise process  $\tilde{v}_k$  ( $\tilde{w}_k$ ) has the same mean and variance as  $v_k$  ( $w_k$ ) but its autocorrelation is a periodic repetition of  $R_v$  ( $R_w$ ) for noise samples within an  $n$ -block.<sup>4</sup> Noise samples from different  $n$ -blocks are independent since the channel is  $n$ -block memoryless. We assume the same power constraint (7) for the  $n$ -CGBC.

The  $n$ -block memoryless Gaussian broadcast channel ( $n$ -MGBC), defined for  $n > m$ , is obtained from the LGBC with memory  $m$  by restricting its outputs. Specifically, the first  $n$ -block of the  $n$ -MGBC derived from the LGBC with finite memory  $m$  is characterized by real input vector  $\{x_k\}_{k=1}^n$  over each  $n$ -block and corresponding real output vectors  $(\hat{y}_{m+1}, \dots, \hat{y}_n)$  and  $(\hat{z}_{m+1}, \dots, \hat{z}_n)$  with  $\hat{y}_k = y_k$  and  $\hat{z}_k = z_k$  defined by (1) and (2), respectively. Previous and subsequent  $n$ -blocks are similarly defined. Note that noise samples from different  $n$ -blocks of the  $n$ -MGBC are independent due to the finite support of  $R_v$  and  $R_w$ . We assume the same power constraint (7) for the  $n$ -MGBC. Note that, as with the circular channel, the  $n$ -MGBC is not  $l$ -block memoryless if  $l$  is not an integer multiple of  $n$ .

### III. DEFINITIONS

In this section, we establish the definitions for a broadcast channel code, error probability, achievable rate vector, and capacity region. Since the LGBC,  $n$ -CGBC, and  $n$ -MGBC all have memory, we cannot use standard definitions for memoryless broadcast channels [6, Sec. 14.6.1], since these

<sup>4</sup>This noise differs from the noise on the circular multiple-access channel in [5], which has  $E[\tilde{v}_k \tilde{v}_l] = R_v[((k-l))_n]$ . This noise is not stationary (and, therefore, does not have a power spectral density), since in general  $R_v[((k-l))_n] \neq R_v[((l-k))_n]$ .

definitions ignore the effect of ISI and noise correlation between codeword transmissions. Specifically, the first transmission of a codeword over the LGBC will cause ISI to the next codeword transmission, and this ISI should be taken into account in the definitions of error probability and achievable rate. Similarly, if an  $l$ -block codeword is transmitted over the  $n$ -CGBC or the  $n$ -MGBC then, for  $l \neq kn$ , this codeword transmission will cause ISI to the next codeword transmission, and this ISI must be taken into account. In addition, the noise samples affecting different codeword transmissions on the LGBC (and the  $n$ -CGBC and  $n$ -MGBC for  $l$ -block codewords with  $l \neq kn$ ) are correlated, and this correlation may impact the optimal decoding strategy and corresponding error probability. Therefore, the noise correlation on finite-memory channels must be taken into account in the definitions of error probability and achievable rate.

There are different ways to incorporate the initial state of a finite memory channel into its capacity definition, depending on what is assumed to be known about this state at the sender and receiver. We follow the approach of [7, Sec. 6], where neither the sender nor the receiver knows this initial state. Thus, the coding and decoding strategies for the channel do not depend on the initial channel state. We also require that a code's error probability go to zero for all possible initial channel states. We show in the next section that, under these definitions, the initial channel state does not impact the channel capacity region for broadcast channels with finite memory. Therefore, we can neglect the impact of ISI and correlated noise between codeword transmissions in the design of the LGBC coding and decoding strategies and in the calculation of the LGBC capacity region.

We now state the definitions for the code, average error probability, achievable rate, and capacity region for broadcast channels with finite memory. Recall from Section II that the initial state of a broadcast channel with finite memory  $m$  is defined as  $\mathcal{S}_0 = (\mathcal{X}_0, \mathcal{V}_0, \mathcal{W}_0)$ , where

$$\begin{aligned}\mathcal{X}_0 &= (x_{-m+1}, \dots, x_0) \\ \mathcal{V}_0 &= (v_{-m+1}, \dots, v_0)\end{aligned}$$

and

$$\mathcal{W}_0 = (w_{-m+1}, \dots, w_0).$$

We let  $\mathcal{X}$  denote the input alphabet of the two-user broadcast channel,  $\mathcal{Y}$  denote the channel output alphabet at the first receiver, and  $\mathcal{Z}$  denote the channel output alphabet at the second receiver. For the LGBC,  $n$ -CGBC, and  $n$ -MGBC,  $\mathcal{X} = \mathcal{Y} = \mathcal{Z} = \mathbb{R}$ .

*Definition:* An  $[(M_1, M_2), l]$  code for a two-user broadcast channel with independent information consists of two sets of integers

$$\mathcal{M}_1 = \{1, \dots, M_1\} \quad \mathcal{M}_2 = \{1, \dots, M_2\}$$

a code mapping

$$X^l: \mathcal{M}_1 \times \mathcal{M}_2 \rightarrow \mathcal{X}^l \quad X^l(W_1, W_2) = (X_1, \dots, X_l)$$

and two decoders

$$d_1: \mathcal{Y}^l \rightarrow \mathcal{M}_1, \quad d_1(Y^l) = \hat{W}_1$$

$$d_2: \mathcal{Z}^l \rightarrow \mathcal{M}_2, \quad d_2(Z^l) = \hat{W}_2.$$

The set of codewords for the broadcast channel is given by

$$\{x^l(w_1, w_2): (w_1, w_2) \in \mathcal{M}_1 \times \mathcal{M}_2\}.$$

The rate pair  $(R_1, R_2)$  associated with an  $[(M_1, M_2), l]$  code is given by

$$\left( R_1 = \frac{1}{l} \log M_1, \quad R_2 = \frac{1}{l} \log M_2 \right),$$

where  $R_1$  and  $R_2$  are the independent information rates sent to receivers 1 and 2, respectively.

*Definition:* The average error probability of an  $[(M_1, M_2), l]$  code for a two-user broadcast channel with finite memory and initial state  $\mathcal{S}_0$ , assuming a uniform message distribution on  $(W_1, W_2) \in \mathcal{M}_1 \times \mathcal{M}_2$ , is defined as

$$\begin{aligned}P_e^l(\mathcal{S}_0) &= \frac{1}{M_1 M_2} \sum_{(w_1, w_2) \in \mathcal{M}_1 \times \mathcal{M}_2} \\ &\Pr [d_1(Y^l) \neq w_1 \vee d_2(Z^l) \neq w_2 | \\ &\quad (W_1 = w_1, W_2 = w_2), \mathcal{S}_0].\end{aligned}$$

Note that for  $n$ -block memoryless channels, including the  $n$ -CGBC and the  $n$ -MGBC, an  $[(M_1, M_2), Nn]$  code for any integer  $N$  has error probability that is independent of the initial channel state  $\mathcal{S}_0$  by (8).

*Definition:* A rate pair  $(R_1, R_2)$  is said to be achievable for a two-user broadcast channel with finite memory if for every  $\gamma > 0$  there exists a sequence of  $[(M_1^{(l)}, M_2^{(l)}), l]$  codes for this channel with  $(1/l) \log M_i^{(l)} \geq R_i - \gamma$  for all  $l$  and  $\sup_{\mathcal{S}_0} P_e^l(\mathcal{S}_0) \rightarrow 0$  as  $l \rightarrow \infty$ .

*Definition:* The capacity region for a broadcast channel is the closure of the set of achievable rates.

We denote the capacity region for the LGBC, obtained using the above definitions, by  $\mathcal{C}$ . The capacity region for the  $n$ -CGBC, denoted by  $\mathcal{C}_n$ , is derived in Section V. Since the  $n$ -CGBC is, by definition, different for every  $n$ ,  $\mathcal{C}_n$  varies with  $n$ . Therefore, the region  $\lim_{n \rightarrow \infty} \mathcal{C}_n$ , if it exists, does not define a capacity region for the CGBC since the  $n$ -CGBC changes with  $n$ . We denote the capacity region of the  $n$ -MGBC by  $\mathcal{Q}_n$ . In general, the  $n$ -MGBC and its corresponding capacity region  $\mathcal{Q}_n$  vary with  $n$ . Therefore, as with the  $n$ -CGBC, the region  $\lim_{n \rightarrow \infty} \mathcal{Q}_n$ , if it exists, does not define a capacity region for the MGBC.

Definitions for the code, average error probability, achievable rate pair, and capacity region for memoryless broadcast channels with common information are given in [6] and [10]. For finite-memory broadcast channels with common information the definitions for the channel code and capacity region are the same as for the memoryless channel. The definitions of average error probability and achievable rate for these channels must be modified to take into account the channel's initial state; these modifications are analogous to the modifications for the finite-memory broadcast channel with independent information described above.

#### IV. ACHIEVABLE RATES AND CAPACITY

In this section we show that the capacity region  $\mathcal{C}$  of the broadcast channel in Fig. 1 is equal to the capacity region of the  $n$ -CGBC and the  $n$ -MGBC as  $n$  goes to infinity. In fact, we prove a more general result. We show that if we extend the definitions of the LGBC, the  $n$ -CGBC, and the  $n$ -MGBC given in Section III to a synchronous Gaussian multiterminal channel [6, Sec. 14.10], then the capacity region of all three of these multiterminal channels is the same in the limit as  $n$  goes to infinity. Since the broadcast channel is a special case of a synchronous multiterminal channel, we get our desired result. Thus, the capacity region  $\mathcal{C}$  of the LGBC can be computed as the limit of  $C_n$ , the capacity region of the  $n$ -CGBC, as  $n$  grows to infinity. This limit is derived in Section V.

Theorem 1 contains the main result for synchronous multiterminal channels. The precise definitions of a finite-memory linear,  $n$ -block circular, and  $n$ -block memoryless synchronous Gaussian multiterminal channel as well as the proof of Theorem 1 are given in Appendix A. A key fact used in the proof of Theorem 1 is that for a channel with impulse response of duration  $m$  and any  $n \geq m$ , linear convolution becomes  $n$ -circular when  $x_k = x_{n+k}$ ,  $k = 1, \dots, m$ , and  $n$ -circular convolution becomes linear when  $x_k = x_{k-n}$ ,  $k = n - m + 1, \dots, n$ . Corollary 1 states the application of Theorem 1 to broadcast channels. Although Theorem 1 (and Corollary 1) assume independent information is sent to each user, these results can be easily extended to the case of common information. Therefore, Corollary 1 applies to broadcast channels with or without common information.

*Theorem 1:* Let  $\mathcal{C}$  denote the capacity region of a linear finite-memory synchronous Gaussian multiterminal channel,  $C_n$  denote the capacity region of an  $n$ -block circular synchronous Gaussian multiterminal channel, and  $Q_n$  denote the capacity region of an  $n$ -block memoryless synchronous Gaussian multiterminal channel, as defined in Appendix A. Then

$$\begin{aligned} \overline{\left(\limsup_{n \rightarrow \infty} C_n\right)} &= \overline{\left(\limsup_{n \rightarrow \infty} Q_n\right)} = \overline{\left(\bigcup_{n > m} Q_n\right)} \\ &= \mathcal{C} = \overline{\left(\liminf_{n \rightarrow \infty} C_n\right)} = \overline{\left(\liminf_{n \rightarrow \infty} Q_n\right)} \end{aligned}$$

where  $\overline{(\cdot)}$  denotes closure.

*Proof:* See Appendix A.

*Corollary 1:* For  $\mathcal{C}$ ,  $C_n$ , and  $Q_n$  as defined in Section III we have that

$$\begin{aligned} \overline{\left(\limsup_{n \rightarrow \infty} C_n\right)} &= \overline{\left(\limsup_{n \rightarrow \infty} Q_n\right)} = \overline{\left(\bigcup_{n > m} Q_n\right)} \\ &= \mathcal{C} = \overline{\left(\liminf_{n \rightarrow \infty} C_n\right)} = \overline{\left(\liminf_{n \rightarrow \infty} Q_n\right)}. \end{aligned}$$

Note that single-user and synchronous multiple-access channels are also special cases of the synchronous multiterminal channel defined in Appendix A. Thus, [3, Corollary 1] and [4, Theorem 1] are also corollaries of Theorem 1, i.e., the finite-memory, circular, and block-memoryless single-user

(synchronous multiple-access) channels all have the same capacity (capacity region) in the limit of infinite block length.

#### V. THE LIMITING CAPACITY REGION OF THE $n$ -CGBC

We have shown in the previous section and Appendix A that the capacity region

$$\mathcal{C} = \overline{\left(\liminf_{n \rightarrow \infty} C_n\right)} = \overline{\left(\limsup_{n \rightarrow \infty} C_n\right)} = \overline{\left(\lim_{n \rightarrow \infty} C_n\right)}$$

where  $C_n$  is the capacity region of the  $n$ -CGBC. We now derive an explicit expression for  $C_n$  and for the limiting region  $\mathcal{C}$ .

For the  $n$ -CGBC,  $n > m$ , consider the input sequence over one  $n$ -block  $\{x_k\}_{k=1}^n$ . The corresponding output sequences  $\{\tilde{y}_k\}_{k=1}^n$  and  $\{\tilde{z}_k\}_{k=1}^n$  have  $\tilde{y}_k$  and  $\tilde{z}_k$  given by (9) and (10), respectively, and are independent of the initial channel state  $\mathcal{S}_0$  by (8). Let  $\{\tilde{h}_i\}_{i=0}^{n-1}$  and  $\{\tilde{g}_i\}_{i=0}^{n-1}$  denote the impulse response of length  $n$  obtained by padding the impulse responses  $\{h_i\}$  and  $\{g_i\}$  with  $n - m - 1$  zeros

$$\{\tilde{h}_i\}_{i=0}^{n-1} = (h_0, \dots, h_m, 0, \dots, 0)$$

and

$$\{\tilde{g}_i\}_{i=0}^{n-1} = (g_0, \dots, g_m, 0, \dots, 0).$$

Then we can rewrite (9) and (10) as

$$\tilde{y}_k = \sum_{i=0}^{n-1} \tilde{h}_i x_{((k-i))_n} + \tilde{v}_k \quad (11)$$

and

$$\tilde{z}_k = \sum_{i=0}^{n-1} \tilde{g}_i x_{((k-i))_n} + \tilde{w}_k. \quad (12)$$

Since invertible operations do not affect capacity, we can apply the DFT to (11) and (12) without affecting the corresponding capacity region. The  $n$ -point DFT  $(B_1, \dots, B_n)$  of the sequence  $(b_1, \dots, b_n)$  is defined as

$$B_k = \sum_{i=1}^n b_i e^{-j2\pi i k/n}. \quad (13)$$

Since (13) is periodic with period  $n$ , the same transform is used for a sequence  $(b_0, \dots, b_{n-1})$  by substituting  $b_0$  for  $b_n$  in (13), and then by periodicity  $B_0 = B_n$ . Applying the DFT to both sides of (11) and (12) we obtain

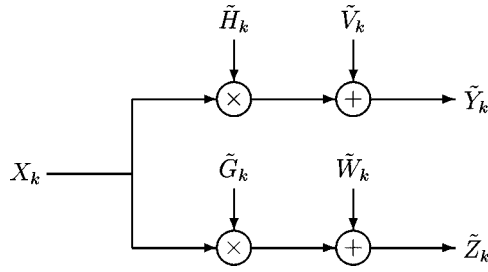
$$\tilde{Y}_k = \tilde{H}_k X_k + \tilde{V}_k \quad (14)$$

and

$$\tilde{Z}_k = \tilde{G}_k X_k + \tilde{W}_k \quad (15)$$

for  $1 \leq k \leq n$ . Thus, the  $n$ -CGBC is equivalent to a set of  $n$  two-user parallel channels with  $k$ th-component channel as shown in Fig. 2.

We will shortly show that if  $\tilde{H}_k \neq 0$  and  $\tilde{G}_k \neq 0$  then the  $k$ th-component channel in Fig. 2 is a degraded broadcast channel. However, it is clear from this figure that if  $\tilde{H}_k \neq 0$  and  $\tilde{G}_k = 0$  then the  $k$ th-component channel reduces to a single-user channel for receiver 1. Similarly, if  $\tilde{H}_k = 0$  and  $\tilde{G}_k \neq 0$  then the  $k$ th-component channel reduces to a single-user channel for receiver 2. Finally, if  $\tilde{H}_k = \tilde{G}_k = 0$  then no information can be transmitted to either user on the

Fig. 2.  $k$ th-component channel.

$k$ th-component channel. Rather than treat each of these cases separately, we will proceed under the assumption that  $\tilde{H}_k \neq 0$  and  $\tilde{G}_k \neq 0$ , then show that the resulting rate region yields the correct rate pair for the  $k$ th-component channel in all cases.

Let us now return to the DFT equations (14) and (15). For any real sequence  $b^n$ , its DFT  $B^n$  has the property that  $B_k = B_{n-k}^*$ ,  $1 \leq k < n$  [11], where  $B^*$  denotes the complex conjugate of  $B$ . Using this property we can reconstruct the entire sequence  $b^n$  from  $B_n$  and  $B^L$  for  $L = \lfloor n/2 \rfloor$ . Thus, we can discard the DFT terms  $B_{L+1}^{n-1}$  without losing any information. Note that  $B_n$  is real for  $n$  even or odd, and  $B_L$  is real for  $n$  even. We define the following transform, which operates on the complex components  $B_n$  and  $B^L$  [3, eq. (24a)]

$$B'_k = \begin{cases} B_k = B_k^R, & k = \begin{cases} n, & n \text{ odd} \\ n \text{ or } L, & n \text{ even} \end{cases} \\ \sqrt{2}B_k^R, & 1 \leq k \leq \begin{cases} L, & n \text{ odd} \\ L-1, & n \text{ even} \end{cases} \\ \sqrt{2}B_{n-k}^I, & L < k < n \end{cases} \quad (16)$$

where  $B_k^R = \text{Re}[B_k]$  and  $B_k^I = \text{Im}[B_k]$ . The transformation (16) is invertible with real outputs and inverse given by [3, eq. (24b)] (with 0 replaced by  $n$ ). We now divide both sides of (14) by  $\tilde{H}_k$  and both sides of (15) by  $\tilde{G}_k$  to get

$$\frac{\tilde{Y}_k}{\tilde{H}_k} = X_k + \frac{\tilde{V}_k}{\tilde{H}_k} \quad (17)$$

and

$$\frac{\tilde{Z}_k}{\tilde{G}_k} = X_k + \frac{\tilde{W}_k}{\tilde{G}_k}. \quad (18)$$

Now applying transform (16) to both sides of (17) and (18) and setting  $Y'_k = (\tilde{Y}_k/\tilde{H}_k)'$ ,  $V'_k = (\tilde{V}_k/\tilde{H}_k)'$ ,  $Z'_k = (\tilde{Z}_k/\tilde{G}_k)'$ , and  $W'_k = (\tilde{W}_k/\tilde{G}_k)'$  yields

$$Y'_k = X'_k + V'_k \quad (19)$$

and

$$Z'_k = X'_k + W'_k. \quad (20)$$

We show in Appendix B that under this transformation with  $n \geq 2m$  the  $V'_k$  and  $W'_k$  are statistically independent real white Gaussian random variables with mean zero and variance

$$\sigma_{V'_k}^2 = \frac{nN_v(2\pi k/n)}{|\tilde{H}_k|^2} \triangleq nN_{1k} \quad (21)$$

and

$$\sigma_{W'_k}^2 = \frac{nN_w(2\pi k/n)}{|\tilde{G}_k|^2} \triangleq nN_{2k}. \quad (22)$$

Moreover, using Parseval's relation for the DFT and the fact that the transform (16) conserves average power yields the following power constraint on  $X'_k$  [3]:

$$\sum_{k=1}^n E[(X'_k)^2] \leq n^2 P. \quad (23)$$

Using the transform described above, the  $n$ -CGBC reduces to the final equivalent channel model shown in Fig. 3. This model consists of  $n$  parallel independent degraded broadcast channels with additive white Gaussian noise, where the noise variances at the two receivers are given by (21) and (22). By [8], [9] we have that the capacity region for this set of parallel degraded broadcast channels, assuming independent information only, is given by the closure of the convex hull of all  $(R_1, R_2)$  satisfying

$$R_1 \leq \sum_{k: N_{1k} \leq N_{2k}} C\left(\frac{\alpha_k P_k}{N_{1k}}\right) + \sum_{k: N_{2k} < N_{1k}} C\left(\frac{\bar{\alpha}_k P_k}{N_{1k} + \alpha_k P_k}\right) \quad (24)$$

$$R_2 \leq \sum_{k: N_{2k} < N_{1k}} C\left(\frac{\alpha_k P_k}{N_{2k}}\right) + \sum_{k: N_{1k} \leq N_{2k}} C\left(\frac{\bar{\alpha}_k P_k}{N_{2k} + \alpha_k P_k}\right) \quad (25)$$

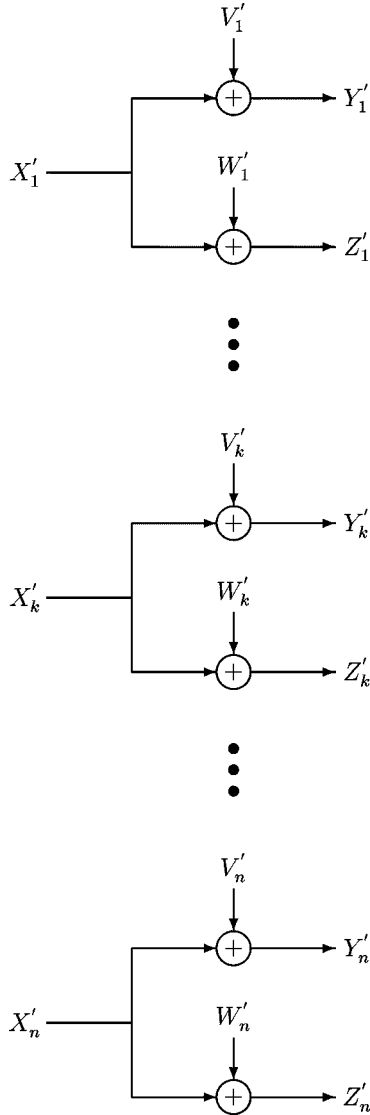
where  $C(x) = 0.5 \log(1+x)$ ,  $\bar{\alpha}_k = 1 - \alpha_k$ , and  $0 \leq \alpha_k \leq 1$  for each  $k$ . In these equations,  $P_k$  is the total power allocated to the  $k$ th-component channel,  $\alpha_k$  is the fraction of  $P_k$  allocated to the user on channel  $k$  with less noise, and  $\bar{\alpha}_k$  is the fraction of  $P_k$  allocated to the user on channel  $k$  with more noise. Equivalently, power  $\bar{\alpha}_k P_k$  is used for the cloud centers of the superposition code designed for the  $k$ th-component degraded broadcast channel, while power  $\alpha_k P_k$  is used for the fine points within the clouds [12]. Note that in (24) and (25) we have scaled both the noise and signal power by  $n$ . Thus, from (23) the set  $\{P_k\}$  must satisfy the power constraint

$$\frac{1}{n} \sum_{k=1}^n P_k \leq P. \quad (26)$$

In Section VI we give numerical examples of the optimal power allocation within and between parallel degraded broadcast channels for several examples.

The capacity region (24) and (25) is derived under the assumption that  $|\tilde{H}_k|$  and  $|\tilde{G}_k|$  are both nonzero. If  $|\tilde{H}_k| = |\tilde{G}_k| = 0$  for some  $k$  then  $N_{1k} = N_{2k} = \infty$ , the contribution to  $R_1$  and  $R_2$  by the  $k$ th-component channel is zero, and so neither user receives any information along the  $k$ th-component channel. If  $|\tilde{H}_k| = 0$  while  $|\tilde{G}_k| > 0$  then  $N_{1k} = \infty > N_{2k}$  so from (24), user 1 does not receive any information over the  $k$ th-component channel. Equivalently, the  $k$ th-component channel reduces to a single-user channel for user 2. Similarly, if  $|\tilde{G}_k| = 0$  while  $|\tilde{H}_k| > 0$  then  $N_{2k} = \infty > N_{1k}$  in which case (25) indicates that user 2 will not receive any information over the  $k$ th-component channel, and the  $k$ th-component channel reduces to a single-user channel for user 1. Thus, the rate region (24) and (25) is valid for any  $|\tilde{H}_k| \geq 0$  and  $|\tilde{G}_k| \geq 0$ .

Let  $\omega_k = 2\pi k/n$ . Then by definition,  $|\tilde{H}_k|^2 = |H(\omega_k)|^2$  and  $|\tilde{G}_k|^2 = |G(\omega_k)|^2$ , where  $H(\omega)$  and  $G(\omega)$  are the channel transfer functions defined in (3) and (4), respectively. Similarly,  $N_{1k} = N_v(\omega_k)|H(\omega_k)|^{-2}$  and  $N_{2k} = N_w(\omega_k)|G(\omega_k)|^{-2}$ .

Fig. 3. Equivalent channel model for the  $n$ -CGBC.

Define  $N_1(\omega_k) = N_{1k}$ ,  $N_2(\omega_k) = N_{2k}$ ,  $P(\omega_k) = P_k$ , and  $\alpha(\omega_k) = \alpha_k$ . Then (24) and (25) become

$$R_1 \leq \sum_{k=1}^n [F_1(\omega_k) + F_2(\omega_k)] \quad (27)$$

$$R_2 \leq \sum_{k=1}^n [F_3(\omega_k) + F_4(\omega_k)] \quad (28)$$

where

$$\begin{aligned} F_1(\omega_k) &= C \left( \frac{\alpha(\omega_k)P(\omega_k)}{N_1(\omega_k)} \right) 1[N_1(\omega_k) \leq N_2(\omega_k)] \\ F_2(\omega_k) &= C \left( \frac{\bar{\alpha}(\omega_k)P(\omega_k)}{N_1(\omega_k) + \alpha(\omega_k)P(\omega_k)} \right) 1[N_2(\omega_k) < N_1(\omega_k)] \\ F_3(\omega_k) &= C \left( \frac{\alpha(\omega_k)P(\omega_k)}{N_2(\omega_k)} \right) 1[N_2(\omega_k) < N_1(\omega_k)] \\ F_4(\omega_k) &= C \left( \frac{\bar{\alpha}(\omega_k)P(\omega_k)}{N_2(\omega_k) + \alpha(\omega_k)P(\omega_k)} \right) 1[N_1(\omega_k) \leq N_2(\omega_k)]. \end{aligned}$$

The power constraint (26) becomes

$$\frac{1}{n} \sum_{k=1}^n P(\omega_k) \leq P. \quad (29)$$

By [13, Theorem 11.33b] if a function  $F$  is bounded and almost everywhere continuous on the interval  $[-\pi, \pi]$  then it is Riemann integrable on the interval. Using this theorem and defining  $d\omega_n = 2\pi/n$  we get that

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n F(\omega_k) &= \lim_{n \rightarrow \infty} \frac{1}{2\pi} \sum_{k=1}^n F(\omega_k) d\omega_n \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} F(\omega) d\omega. \end{aligned} \quad (30)$$

The functions  $F_1$ ,  $F_2$ ,  $F_3$ , and  $F_4$  are all bounded since  $C(\cdot)$  is bounded. Moreover, since  $C(\cdot)$  is continuous we need only show that  $1[N_1(\omega) \leq N_2(\omega)]$  and  $1[N_2(\omega) < N_1(\omega)]$  are almost everywhere continuous to apply (30) to (27) and (28), where  $N_1(\omega) = N_v(\omega)|H(\omega)|^{-2}$  and  $N_2(\omega) = N_w(\omega)|G(\omega)|^{-2}$ . But  $|H(\omega)|$  and  $|G(\omega)|$  are, by (3) and (4), continuous functions with a finite number of maxima and minima in any given interval. Similarly, by (5) and (6),  $N_v(\omega)$  and  $N_w(\omega)$  are also continuous functions with a finite number of maxima and minima in any given interval. Thus  $1[N_1(\omega) \leq N_2(\omega)]$  and  $1[N_2(\omega) < N_1(\omega)]$  have a finite number of discontinuities on any given interval, and are therefore almost everywhere continuous.

We now divide (27) and (28) by  $n$ , take the limit as  $n \rightarrow \infty$ , and apply (30) to get that the capacity region  $\mathcal{C}$  is the closure of the convex hull of all  $(R_1, R_2)$  satisfying

$$\begin{aligned} R_1 &\leq \frac{1}{2\pi} \int_{-\pi}^{\pi} C \left( \frac{\alpha(\omega)P(\omega)}{N_1(\omega)} \right) 1[N_1(\omega) \leq N_2(\omega)] d\omega \\ &\quad + \frac{1}{2\pi} \int_{-\pi}^{\pi} C \left( \frac{\bar{\alpha}(\omega)P(\omega)}{N_1(\omega) + \alpha(\omega)P(\omega)} \right) \\ &\quad \times 1[N_2(\omega) < N_1(\omega)] d\omega \\ R_2 &\leq \frac{1}{2\pi} \int_{-\pi}^{\pi} C \left( \frac{\alpha(\omega)P(\omega)}{N_2(\omega)} \right) 1[N_2(\omega) < N_1(\omega)] d\omega \\ &\quad + \frac{1}{2\pi} \int_{-\pi}^{\pi} C \left( \frac{\bar{\alpha}(\omega)P(\omega)}{N_2(\omega) + \alpha(\omega)P(\omega)} \right) \\ &\quad \times 1[N_1(\omega) \leq N_2(\omega)] d\omega \end{aligned}$$

where  $0 \leq \alpha(\omega) \leq 1$ . The same procedure applied to (29) yields the power constraint

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} P(\omega) d\omega \leq P. \quad (31)$$

Note that since  $H(\omega)$ ,  $G(\omega)$ ,  $N_v(\omega)$ , and  $N_w(\omega)$  are all symmetric functions,  $N_1(\omega)$  and  $N_2(\omega)$  are also symmetric, and therefore the capacity and corresponding optimal power allocation need only be calculated on the interval  $[0, \pi]$ . We can thus use this symmetry to calculate the capacity region on the interval  $[0, \pi]$  only. Define

$$\Omega_1 = \{\omega \in [0, \pi]: N_1(\omega) \leq N_2(\omega)\}$$

and

$$\Omega_2 = \{\omega \in [0, \pi]: N_2(\omega) < N_1(\omega)\}.$$

Substituting in for  $C(\cdot)$  and using symmetry we get that the capacity region  $\mathcal{C}$  equals the closure of the convex hull of all points  $(R_1, R_2)$  satisfying

$$R_1 \leq \frac{1}{2\pi} \int_{\Omega_1} \log \left( 1 + \frac{\alpha(\omega)P(\omega)}{N_1(\omega)} \right) d\omega \\ + \frac{1}{2\pi} \int_{\Omega_2} \log \left( 1 + \frac{\bar{\alpha}(\omega)P(\omega)}{N_1(\omega) + \alpha(\omega)P(\omega)} \right) d\omega \quad (32)$$

$$R_2 \leq \frac{1}{2\pi} \int_{\Omega_2} \log \left( 1 + \frac{\alpha(\omega)P(\omega)}{N_2(\omega)} \right) d\omega \\ + \frac{1}{2\pi} \int_{\Omega_1} \log \left( 1 + \frac{\bar{\alpha}(\omega)P(\omega)}{N_2(\omega) + \alpha(\omega)P(\omega)} \right) d\omega \quad (33)$$

where  $0 \leq \alpha(\omega) \leq 1$  and the power constraint (31) becomes

$$\frac{1}{\pi} \int_0^\pi P(\omega) d\omega \leq P. \quad (34)$$

Equations (32) and (33) coupled with the power constraint (34) complete our characterization of the capacity region for broadcast channels with finite ISI and noise correlation, assuming independent information and an average power constraint  $P$ .

We can easily extend the derivation given above to obtain the capacity region when the transmitter sends both common and independent information. The resulting capacity region, given in Appendix C, is characterized implicitly as the intersection of six regions. Because of this implicit characterization it is much harder to derive numerical results for the capacity region with both common and independent information than for the case of independent information only. Our numerical results in the next section focus on the capacity region and optimal power allocation for independent information only.

## VI. ANALYSIS AND EXAMPLES

Numerical examples are useful for better understanding the results presented in Section V. Consider a broadcast channel with two receivers. Let  $H_i(\omega)$  denote the transfer function over the channel associated with receiver  $i$ ,  $i = 1, 2$ . We are interested in finding the set  $\mathcal{C}(H_1, H_2)$  of achievable rate pairs  $(R_1, R_2)$  across the broadcast channel  $(H_1, H_2)$ . Since the achievable rate region is convex (by a time-sharing argument) and closed (by definition), that region is entirely characterized by its support functional  $J((\lambda_1, \lambda_2), (H_1, H_2))$  [14, p. 135] over all  $\lambda_1, \lambda_2 \geq 0$  such that  $\lambda_1 + \lambda_2 = 1$ , where

$$J((\lambda_1, \lambda_2), (H_1, H_2)) = \sup_{(R_1, R_2) \in \mathcal{C}(H_1, H_2)} \sum_{i=1}^2 \lambda_i R_i. \quad (35)$$

In this equation,  $(\lambda_1, \lambda_2)$  describes a prioritization over the rates  $(R_1, R_2)$ . For example,  $\lambda_1 < \lambda_2$  implies that the rate  $R_2$  is of higher priority than the rate  $R_1$ . The vector  $(\lambda_1, \lambda_2)$  is the Lagrangian constant in a convex optimization problem. This vector may be interpreted as the “slope” or direction of a plane tangent to the achievable rate region at a single point. By considering all possible slopes and their associated tangent points, we trace out the convex hull of the achievable rate region. Since the relative rather than the absolute values of  $(\lambda_1, \lambda_2)$  determine its direc-

tion, there is no loss of generality associated with the restriction  $\lambda_1 + \lambda_2 = 1$ .

Combining (35) with (32) and (33), finding the outer boundary of the achievable rate region is equivalent to finding, for all  $\lambda_1, \lambda_2 \geq 0$  such that  $\lambda_1 + \lambda_2 = 1$

$$J((\lambda_1, \lambda_2), (H_1, H_2)) \\ = \max_{\alpha(\omega), P(\omega)} \lambda_1 \left[ \frac{1}{2\pi} \int_{\Omega_1} \log \left( 1 + \frac{\alpha(\omega)P(\omega)}{N_1(\omega)} \right) d\omega \right. \\ \left. + \frac{1}{2\pi} \int_{\Omega_2} \log \left( 1 + \frac{\bar{\alpha}(\omega)P(\omega)}{N_1(\omega) + \alpha(\omega)P(\omega)} \right) d\omega \right] \\ + \lambda_2 \left[ \frac{1}{2\pi} \int_{\Omega_2} \log \left( 1 + \frac{\alpha(\omega)P(\omega)}{N_2(\omega)} \right) d\omega \right. \\ \left. + \frac{1}{2\pi} \int_{\Omega_1} \log \left( 1 + \frac{\bar{\alpha}(\omega)P(\omega)}{N_2(\omega) + \alpha(\omega)P(\omega)} \right) d\omega \right]$$

subject to the constraints

$$\pi^{-1} \int_0^\pi P(\omega) d\omega = P$$

and  $0 \leq \alpha(\omega) \leq 1$  for all  $\omega \in [0, \pi]$ .

Finding  $J((\lambda_1, \lambda_2), (H_1, H_2))$  for a given  $(\lambda_1, \lambda_2)$  requires an optimization over all possible power allocations  $P(\omega)$  over the range of frequencies  $\omega \in [0, \pi]$  plus a power allocation  $\alpha(\omega)$  between the cloud centers  $U$  and the detail points  $X$  sent over a particular channel, where for each  $\omega \in [0, \pi]$ ,  $\bar{\alpha}(\omega)$  is the fraction of power  $P(\omega)$  dedicated to the cloud centers and  $\alpha(\omega) = 1 - \bar{\alpha}(\omega)$  is the remaining portion. Lagrangian optimization may be used to simultaneously optimize over  $\alpha(\omega)$  and  $P(\omega)$  subject to the constraints  $\alpha(\omega) \in [0, 1]$  for all  $\omega$  and  $\pi^{-1} \int_0^\pi P(\omega) d\omega = P$ . Solution of this Lagrangian and consideration of the boundary conditions  $P(\omega) \geq 0$  and  $\alpha(\omega) \in [0, 1]$  for all  $\omega$  yields an optimal interfrequency power allocation  $P(\omega)$  and an optimal intrafrequency power allocation  $\alpha(\omega)$ .

The optimal interfrequency power allocation  $P(\omega)$  is a water-filling solution with two water levels scaled by the priorities  $\lambda_1, \lambda_2$  [8]. That is,

$$P(\omega) = \begin{cases} (\lambda_1 c - N_1(\omega))^+ & \text{if } \lambda_2 c - N_2(\omega) < \lambda_1 c - N_1(\omega) \\ (\lambda_2 c - N_2(\omega))^+ & \text{if } \lambda_2 c - N_2(\omega) \geq \lambda_1 c - N_1(\omega) \end{cases} \quad (36)$$

where  $(x)^+ = x \cdot 1[x > 0]$  and  $c$  is chosen to satisfy the power constraint  $\pi^{-1} \int_0^\pi P(\omega) d\omega = P$  across the frequencies.

Fig. 4 shows the optimal interfrequency power allocation for a channel with additive white Gaussian noise  $[N_v(\omega) = N_w(\omega) = 1 \text{ for all } \omega]$  and a pair of receivers described by one low pass filter and one high pass filter. Results for a variety of  $(\lambda_1, \lambda_2)$  values are given. A similar diagram for a pair of low pass filters appears in Fig. 5. In each graph, the solid lines indicate the noise functions  $N_i(\omega) = |H_i(\omega)|^{-2}$ ,  $i = 1, 2$ , and the height of the shaded region for each frequency describes the power allocated to transmitting information at that frequency. As indicated by (36), at each frequency, water-filling is done on either  $N_1(\omega)$  or  $N_2(\omega)$ . Water level  $\lambda_1 c$  is used at all frequencies  $\omega$  for which water-filling is done on  $N_1(\omega)$  and water level  $\lambda_2 c$  is used at all frequencies  $\omega$  for which water-filling is done on  $N_2(\omega)$ . The water-filling at a



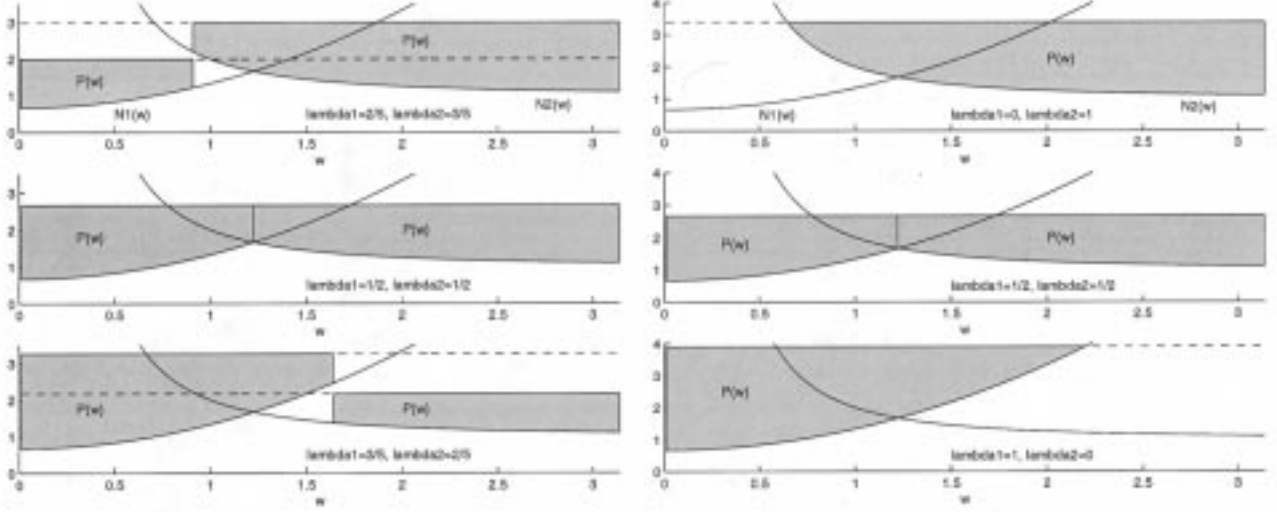


Fig. 4. Optimal power allocation for  $N_1(\omega) = 2(\omega^2 + 1)/3$  (corresponding to a channel characterized by a low-pass filter) and  $N_2(\omega) = (\omega^2 + 1)/\omega^2$  (corresponding to a channel characterized by a high-pass filter). Solid lines indicate the noise functions  $N_i(\omega)$ ,  $i = 1, 2$ . The height of the shaded region for each frequency indicates the power  $P(\omega)$  allocated to transmitting information at that frequency. Since the noise functions are even symmetric, only the frequencies between 0 and  $\pi$  are shown.

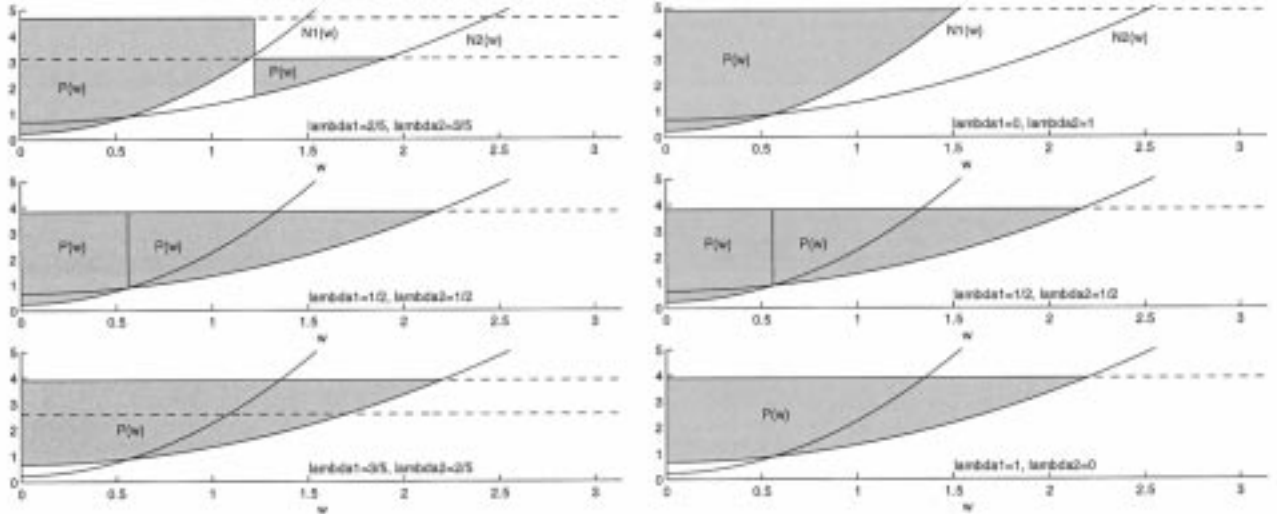


Fig. 5. Optimal power allocation for  $N_1(\omega) = 2(\omega^2 + 1)/3$  and  $N_2(\omega) = 2(\omega^2 + 1)/8$  (both of which correspond to channels characterized by low-pass filters). Solid lines indicate the noise functions  $N_i(\omega)$ ,  $i = 1, 2$ . The height of the shaded region for each frequency indicates the power  $P(\omega)$  allocated to transmitting information at that frequency. Since the noise functions are even symmetric, only the frequencies between 0 and  $\pi$  are shown.

particular frequency  $\omega$  is done on whichever  $N_i(\omega)$ ,  $i = 1, 2$  gives the higher power allocation at that frequency. That is, if

$$\lambda_2 c - N_2(\omega) \geq \lambda_1 c - N_1(\omega)$$

then water-filling is done on  $N_2(\omega)$  whereas if

$$\lambda_2 c - N_2(\omega) < \lambda_1 c - N_1(\omega)$$

then water-filling is done on  $N_1(\omega)$ . No power is allocated at any frequency  $\omega$  such that  $N_1(\omega) > \lambda_1 c$  and  $N_2(\omega) > \lambda_2 c$ . Thus the power allocated at frequency  $\omega$  depends on the priority values described by  $\lambda_1$  and  $\lambda_2$  and the noise levels  $N_1(\omega)$  and  $N_2(\omega)$  at the two receivers.

One interesting consequence of the optimal interfrequency power allocation equation is that the power allocation at a frequency  $\omega$  may be based on the *larger* of the two noise levels. For example, if  $N_1(\omega) > N_2(\omega)$  but  $\lambda_1 c - N_1(\omega) > \lambda_2 c - N_2(\omega)$ , then the power  $P(\omega)$  will be a water-filling on the higher noise  $N_1(\omega)$  (this is possible only if the priority  $\lambda_1$  on rate  $R_1$  is greater than the priority  $\lambda_2$  on rate  $R_2$ ).

The fraction  $\alpha(\omega)$  of the power in a given channel used to send information to the better receiver for that channel likewise varies as a function of the noise functions  $N_i(\omega)$  at that frequency as well as the priorities on the two receivers, as is shown in the equation at the bottom of the following page.

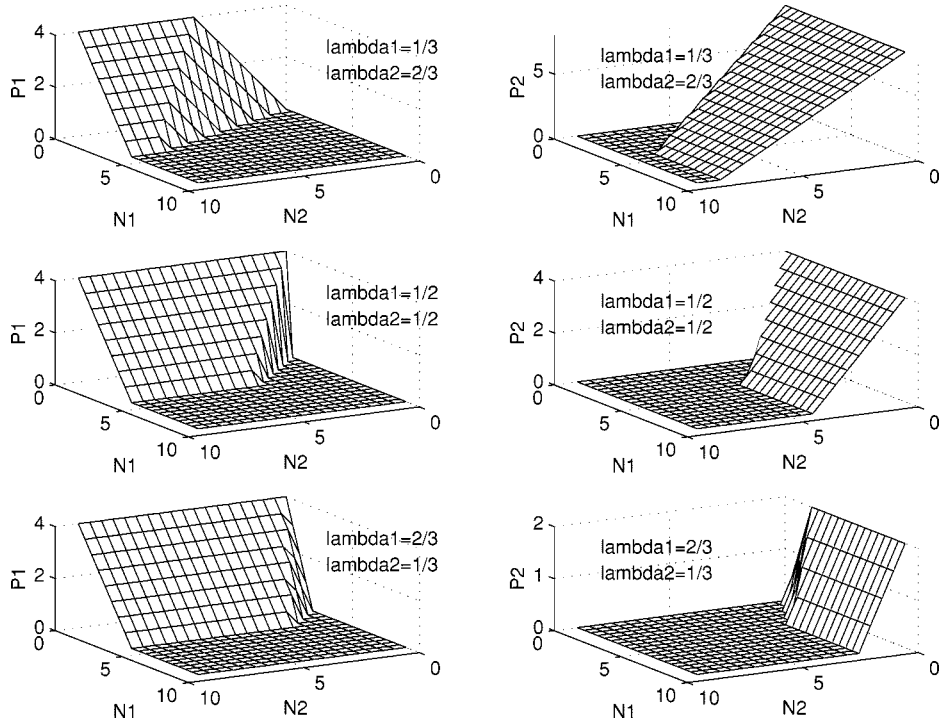


Fig. 6. Optimal power allocation  $P_i(\omega)$  for a single frequency  $\omega$  as a function of  $N_1(\omega)$  and  $N_2(\omega)$ . The power  $P_i(\omega)$ ,  $i = 1, 2$  is the power used to send information to receiver  $i$  at frequency  $\omega$ .

Let  $P_i(\omega)$ ,  $i = 1, 2$  be the power used in sending information to receiver  $i$  at frequency  $\omega$ . Then  $P_1(\omega) + P_2(\omega) = P(\omega)$  for all  $\omega \in [0, \pi]$  and

$$P_i(\omega) = \begin{cases} \alpha(\omega)P(\omega), & \text{if } \omega \in \Omega_i \\ (1 - \alpha(\omega))P(\omega), & \text{if } \omega \in \Omega_i^c \end{cases} \quad i = 1, 2$$

where

$$\Omega_1 = \{\omega \in [0, \pi]: N_1(\omega) \leq N_2(\omega)\}$$

and

$$\Omega_2 = \Omega_1^c = \{\omega \in [0, \pi]: N_2(\omega) < N_1(\omega)\}$$

as defined in Section V-A plot of the optimal power allocations  $P_1$  and  $P_2$  for a single frequency as a function of both  $N_1$  and  $N_2$  at that frequency appears in Fig. 6. Notice that the power is shared between the two channels ( $\alpha(\omega) \in (0, 1)$ ) only when  $\lambda_2/\lambda_1$  falls between  $(P(\omega) + N_2(\omega))/(P(\omega) + N_1(\omega))$  and  $N_2(\omega)/N_1(\omega)$ . All other frequencies are dedicated to sending power to one or the other receiver but not both.

Fig. 7 reworks the power allocation example of Fig. 4 to show the power allocated to each receiver. Dark shading indicates the power allocated to receiver 1 and light shading indicates the power allocated to receiver 2. Similar results appear

in Fig. 8 for the example of Fig. 5. The dotted vertical line in each graph indicates the location of the boundary between the set  $\{\omega: \lambda_2 c - N_2(\omega) < \lambda_1 c - N_1(\omega)\}$  on which water-filling is done on  $N_1(\omega)$  and the set  $\{\omega: \lambda_2 c - N_2(\omega) \geq \lambda_1 c - N_1(\omega)\}$  on which water-filling is done on  $N_2(\omega)$ . Notice that while the interchannel power allocation does its water-filling on a single filter for each frequency,  $P(\omega) = \lambda_1 c - N_1(\omega)$  does not imply that all of the power is being allocated to user 1, nor does  $P(\omega) = \lambda_2 c - N_2(\omega)$  imply that all of the power is being allocated to user 2. In fact, at any single frequency, power sharing between the two users occurs only if  $\lambda_2/\lambda_1$  falls between  $(P(\omega) + N_2(\omega))/(P(\omega) + N_1(\omega))$  and  $N_2(\omega)/N_1(\omega)$ , as discussed earlier. One implication of this equation is that power sharing only occurs at values of  $\omega$  for which water-filling is done on the channel with higher noise. Another implication is that when the gap between  $(P(\omega) + N_2(\omega))/(P(\omega) + N_1(\omega))$  and  $N_2(\omega)/N_1(\omega)$  is small, power sharing is done at very few frequencies. In this case, frequency division, where each frequency is assigned to just one of the two users, will achieve near-capacity rates.

Given the optimal power allocation  $\alpha(\omega)$  between receivers at a given frequency and the optimal power allocation  $P(\omega)$

$$\alpha(\omega)P(\omega) = \begin{cases} 0 & 1 < \frac{N_2(\omega)}{N_1(\omega)} \leq \frac{\lambda_2}{\lambda_1} & \vee & \frac{\lambda_2}{\lambda_1} \leq \frac{N_2(\omega)}{N_1(\omega)} \leq 1 \\ \frac{\lambda_1 N_2(\omega) - \lambda_2 N_1(\omega)}{\lambda_2 - \lambda_1} & 1 < \frac{P(\omega) + N_2(\omega)}{P(\omega) + N_1(\omega)} \leq \frac{\lambda_2}{\lambda_1} < \frac{N_2(\omega)}{N_1(\omega)} & \vee & \frac{N_2(\omega)}{N_1(\omega)} < \frac{\lambda_2}{\lambda_1} \leq \frac{P(\omega) + N_2(\omega)}{P(\omega) + N_1(\omega)} < 1 \\ P(\omega) & \frac{\lambda_2}{\lambda_1} < \frac{P(\omega) + N_2(\omega)}{P(\omega) + N_1(\omega)} < \frac{N_2(\omega)}{N_1(\omega)} & \vee & \frac{N_2(\omega)}{N_1(\omega)} < \frac{P(\omega) + N_2(\omega)}{P(\omega) + N_1(\omega)} < \frac{\lambda_2}{\lambda_1} \end{cases}$$

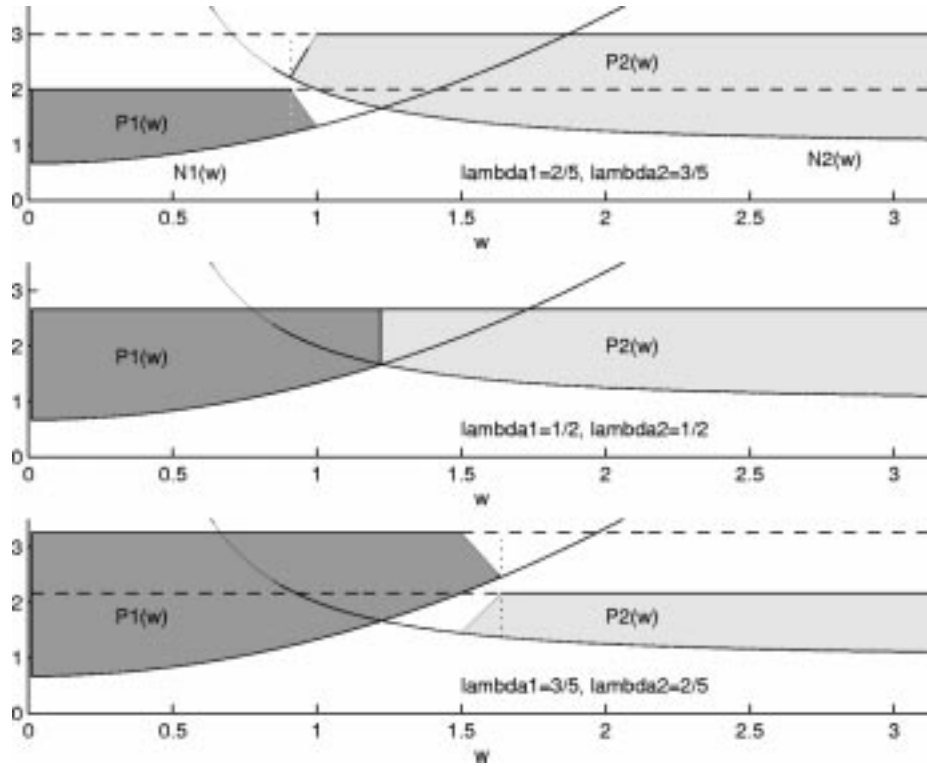


Fig. 7. Optimal power allocated to each receiver for  $N_1(\omega) = 2(\omega^2 + 1)/3$  and  $N_2(\omega) = (\omega^2 + 1)/\omega^2$ . Solid lines indicate the noise functions  $N_i(\omega)$ ,  $i = 1, 2$ . The height of the shaded region for each frequency indicates the power allocated to transmitting information at that frequency. Dark shading indicates power transmitted to receiver 1. Light shading indicates power transmitted to receiver 2. The dotted vertical line in each graph marks the location of the boundary between the set  $\{\omega: \lambda_2 c - N_2(\omega) < \lambda_1 c - N_1(\omega)\}$  on which water-filling is done on  $N_1(\omega)$  and the set  $\{\omega: \lambda_2 c - N_2(\omega) \geq \lambda_1 c - N_1(\omega)\}$  on which water-filling is done on  $N_2(\omega)$ .

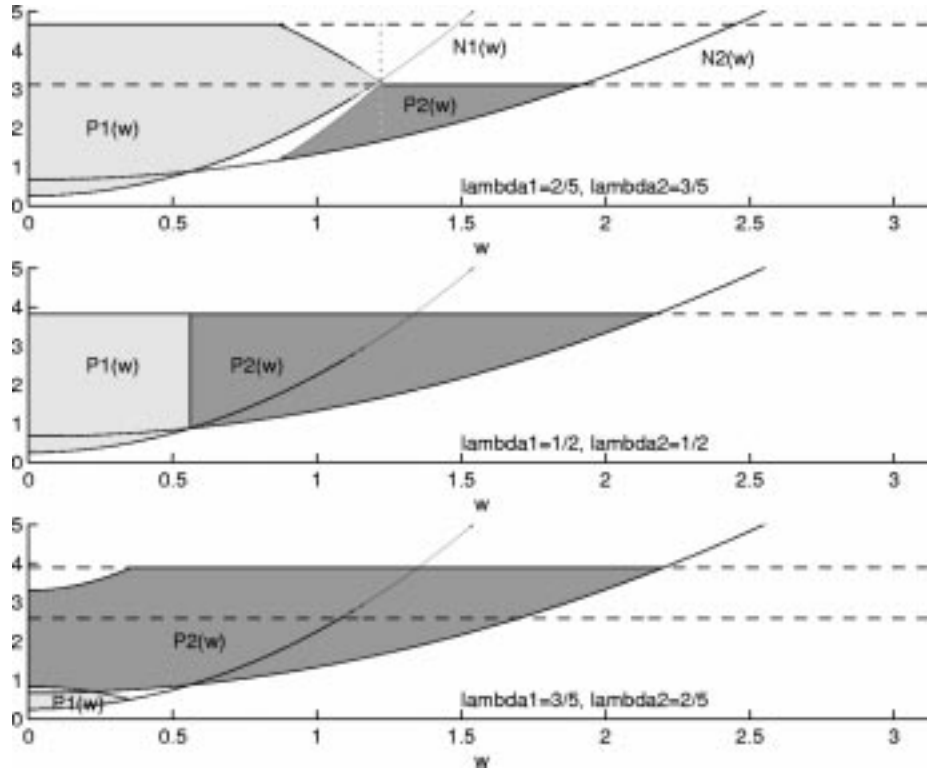


Fig. 8. Optimal power allocated to each receiver for  $N_1(\omega) = 2(\omega^2 + 1)/3$  and  $N_2(\omega) = 2(\omega^2 + 1/8)$ . Solid lines indicate the noise functions  $N_i(\omega)$ ,  $i = 1, 2$ . The height of the shaded region for each frequency indicates the power allocated to transmitting information at that frequency. Dark shading indicates power transmitted to receiver 1. Light shading indicates power transmitted to receiver 2. The dotted vertical line in each graph marks the location of the boundary between the set  $\{\omega: \lambda_2 c - N_2(\omega) < \lambda_1 c - N_1(\omega)\}$  on which water-filling is done on  $N_1(\omega)$  and the set  $\{\omega: \lambda_2 c - N_2(\omega) \geq \lambda_1 c - N_1(\omega)\}$  on which water-filling is done on  $N_2(\omega)$ .

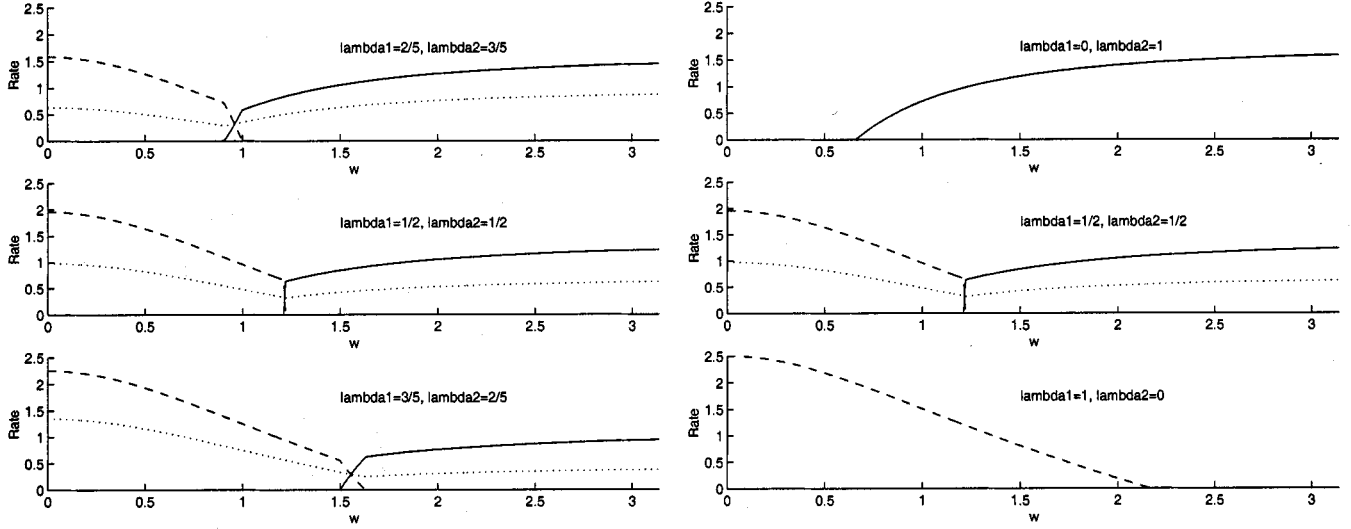


Fig. 9. Rate functions  $r_1(\omega)$  and  $r_2(\omega)$  for  $N_1(\omega) = 2(\omega^2 + 1)/3$  and  $N_2(\omega) = (\omega^2 + 1)/\omega^2$ . Dashed lines indicate rate  $r_1(\omega)$  while solid lines indicate rate  $r_2(\omega)$ . The marginal Lagrangian function  $j(\omega) = \lambda_1 r_1(\omega) + \lambda_2 r_2(\omega)$  appears as a dotted line.

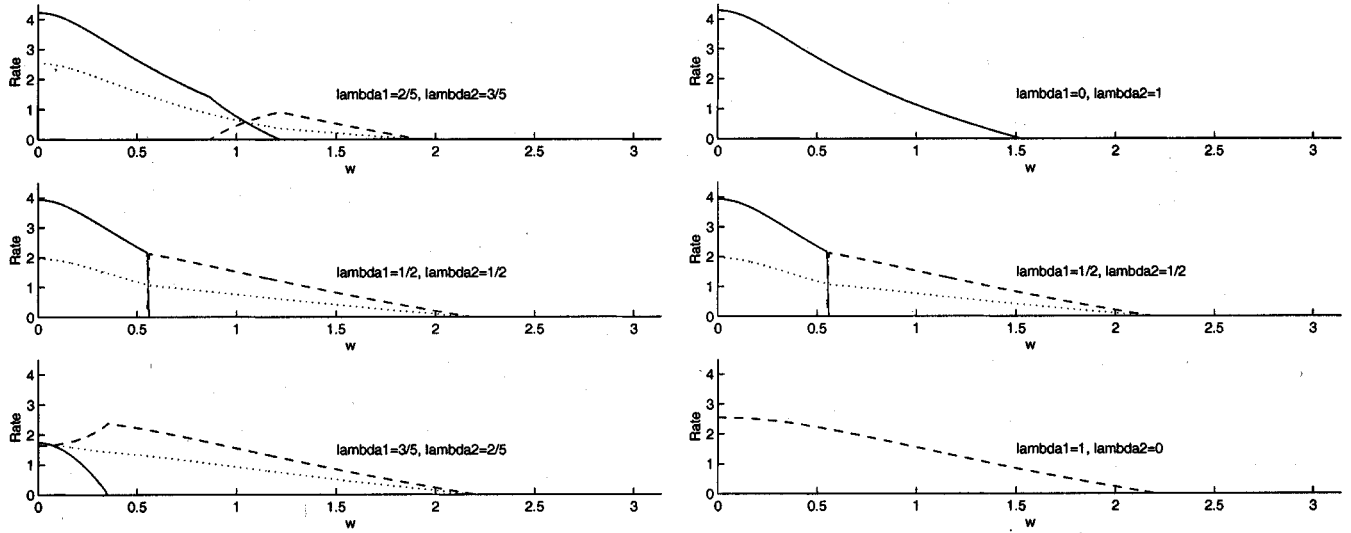


Fig. 10. Rate functions  $r_1(\omega)$  and  $r_2(\omega)$  for  $N_1(\omega) = 2(\omega^2 + 1)/3$  and  $N_2(\omega) = 2(\omega^2 + 1/8)$ . Dashed lines indicate rate  $r_1(\omega)$  while solid lines indicate rate  $r_2(\omega)$ . The marginal Lagrangian function  $j(\omega) = \lambda_1 r_1(\omega) + \lambda_2 r_2(\omega)$  appears as a dotted line.

between frequencies, we can now plot the rate  $r_i(\omega)$ ,  $i = 1, 2$  sent to receiver  $i$  at frequency  $\omega$ , where

$$r_i(\omega) = \begin{cases} \frac{1}{2} \log \left( 1 + \frac{\alpha(\omega)P(\omega)}{N_i(\omega)} \right), & \text{if } \omega \in \Omega_i \\ \frac{1}{2} \log \left( 1 + \frac{\bar{\alpha}(\omega)P(\omega)}{N_i(\omega) + \alpha(\omega)P(\omega)} \right), & \text{if } \omega \in \Omega_i^c \end{cases}$$

for  $i = 1, 2$

is the marginal portion of rate  $R_i$  contributed by frequency  $\omega$  in (32) and (33). A plot of  $r_1(\omega)$  (dashed line) and  $r_2(\omega)$  (solid line) for the example of Fig. 4 appears in Fig. 9. Also included in this figure is a plot of  $j(\omega) = \lambda_1 r_1(\omega) + \lambda_2 r_2(\omega)$  (dotted line).

A plot of these marginal rates for the pair of low-pass filters first described in Fig. 5 appears in Fig. 10

Each  $r_i(\omega)$  curve in the previous figures corresponds to a single value of the priority vector  $(\lambda_1, \lambda_2)$ . Integrating any  $r_1(\omega)$  and  $r_2(\omega)$  over all  $\omega \in [0, \pi]$  gives the optimal rate  $(R_1, R_2)$  associated with that Lagrangian constant. Repeating this process for each  $(\lambda_1, \lambda_2)$  such that  $\lambda_1 + \lambda_2 = 1$  traces out the capacity region associated with the given channel without common information. The capacity regions for the two examples considered above are shown in Fig. 11.

## VII. CONCLUSION

We have obtained the capacity region of a finite-memory broadcast channel with ISI and colored Gaussian noise under

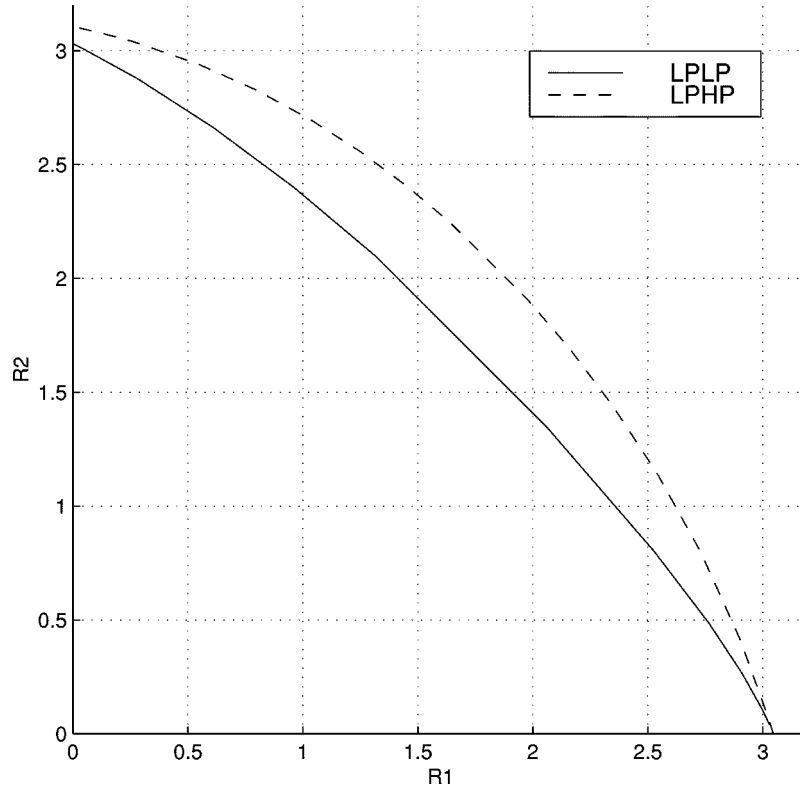


Fig. 11. The outer convex hulls of the capacity regions of two different broadcast channels. The first example, labeled LPHP shows the capacity region of the broadcast channel described by  $N_1(\omega) = |H_1(\omega)|^{-2} = 2(\omega^2 + 1)/3$  and  $N_2(\omega) = |H_2(\omega)|^{-2} = (\omega^2 + 1)/\omega^2$ . The second example, labeled LPLP, shows the capacity region of the broadcast channel described by  $N_1(\omega) = 2(\omega^2 + 1)/3$  and  $N_2(\omega) = 2(\omega^2 + 1/8)$ .

an average input power constraint. This capacity region is shown to equal the capacity region of an  $n$ -circular Gaussian broadcast channel as  $n$  grows to infinity. The  $n$ -circular channel for any  $n$  can be decomposed using the DFT into a set of parallel independent degraded broadcast channels, for which the capacity is known. Taking the limit of this capacity region as  $n$  increases to infinity yields the desired capacity region.

We then solve numerically for the convex hull of the capacity region and the corresponding optimal power allocation for two example channels. The optimal power allocation is obtained via a water-filling in the frequency domain with two different water levels, where the water levels depend on the relative priorities of the different users. We also describe the frequencies that have power sharing between the users. Numerical results for the power allocation and capacity are obtained for both examples, where the first example has one low-pass channel and one high-pass channel, and the second example has two low-pass channels. The graphical interpretations for power allocation and capacity give interesting insights into the protocol for broadcast channels sending independent information. In particular, for broadcast channels where there is just a small range of frequencies with true power sharing, frequency division of the broadcast spectrum with optimal power allocation achieves close to channel capacity.

Our numerical results are for broadcast channels used to send independent information to multiple users. For broadcast channels with common information the capacity region is implicitly

obtained as the intersection of six regions. Due to this implicit characterization it is difficult to obtain numerical results or any intuition about the optimal power allocation and corresponding capacity region for finite-memory broadcast channels with common information.

#### APPENDIX A

In this appendix, we extend the definitions of the finite-memory, circular, and block-memoryless broadcast channels given in Section III to general synchronous multiterminal channels (networks) [6, Sec. 14.10]. We then prove Theorem 1: that the capacity region of all these channels is the same in the limit of infinite block length.

Consider a discrete-time synchronous multiterminal channel. The channel model consists of  $L$  nodes, where each node potentially sends information to and receives information from all other nodes. Node  $t$  transmits real input sequence

$$\{x_{(t)}[k]\} = (\dots, x_{(t)}[1], x_{(t)}[2], \dots)$$

and receives real output sequence

$$\{y_{(t)}[k]\} = (\dots, y_{(t)}[1], y_{(t)}[2], \dots).$$

The synchronous multiterminal channel under consideration is a linear channel with stationary colored additive Gaussian

noise of finite memory. In particular, at time  $k$ , node  $r$  receives channel output  $y_{(r)}[k]$  given by

$$\begin{aligned} y_{(r)}[k] &= \sum_{t=1}^L \left[ \sum_{i=0}^m h_{(t,r)}[i] x_{(t)}[k-i] \right] + V_{(r)}[k] \\ &= \sum_{t=1}^L h_{(t,r)}[k] * x_{(t)}[k] + V_{(r)}[k] \end{aligned}$$

where  $*$  denotes linear convolution. For any node  $r \in \{1, \dots, L\}$ ,  $V_{(r)}[k]$  represents a stationary random receiver noise process with mean zero and autocorrelation function  $R_{(r)}[i]$ . By assumption, the autocorrelation function has finite support, and thus there exists a finite  $i_{\max}$  such that  $R_{(r)}[i] = 0$  for all  $|i| \geq i_{\max}$  and all  $r$ . For each pair of nodes  $(t, r) \in \{1, \dots, L\}^2$ ,  $\{h_{(t,r)}[k]\}_{k=0}^m$  denotes the real finite impulse response of the channel through which the signal transmitted by node  $t$  and received by node  $r$  travels. The duration  $m$  of that impulse response does not vary as a function of  $(t, r)$ , although for some  $(t, r)$  pairs setting  $h_{(t,r)}[k] = 0$  for  $k < m$  will effectively create a shorter impulse response for these channels. We call  $m$  the *channel memory* and, without loss of generality, assume that  $m \geq i_{\max}$ .<sup>5</sup> The channels have ISI since the channel output at time  $k$  depends on the input symbols  $\{x_{(t)}[k]\}_{t \in \{1, \dots, L\}}$  at time  $k$  as well as previous input symbols  $\{x_{(t)}[i]\}_{t \in \{1, \dots, L\}, i < k}$ . In addition, the noise sample  $V_{(r)}[k]$  at time  $k$  is correlated with noise samples  $V_{(r)}[i]$  at times  $i < k$ .

While the above channel model allows all nodes to simultaneously transmit and receive information, notice that it does not preclude channel scenarios where some nodes act exclusively as transmitters and others act exclusively as receivers. These special cases may be achieved by setting appropriate filter coefficients to zero. For example,  $h_{(t,\cdot)}[k]$  equals zero for all  $k \in \{0, \dots, m\}$  when index  $t$  designates a node that transmits no information while  $h_{(\cdot,r)}[k]$  equals zero for all  $k \in \{0, \dots, m\}$  when index  $r$  designates a node that receives no information. Further,  $h_{(t,t)}[k]$  generally equals zero for all  $k \in \{0, \dots, m\}$  and all  $t \in \{1, \dots, L\}$  since there is no need for any node to transmit information to itself. For example, a broadcast system with a single transmitter (node 1) and two receivers (nodes 2 and 3) may be achieved in an  $L = 3$  system by setting  $h_{(t,r)}[k] = 0$  for all  $t \in \{2, 3\}$ ,  $r \in \{1, 2, 3\}$ , and  $k \in \{0, \dots, m\}$  and for  $t = 1, r = 1$ , and  $k \in \{0, \dots, m\}$ , thereby specifying that only node 1 transmits information and nodes 2 and 3 receive this information. Specifically, nodes 2 and 3 receive

$$\begin{aligned} y_{(2)}[k] &= \sum_{i=0}^m h_{(1,2)}[i] x_{(1)}[k-i] + V_{(2)}[k] \\ &= h_{(1,2)}[k] * x_{(1)}[k] + V_{(2)}[k], \\ y_{(3)}[k] &= \sum_{i=0}^m h_{(1,3)}[i] x_{(1)}[k-i] + V_{(3)}[k] \\ &= h_{(1,3)}[k] * x_{(1)}[k] + V_{(3)}[k] \end{aligned}$$

<sup>5</sup>If  $m < i_{\max}$ , we pad all impulse responses with zeros to obtain the desired property.

while node 1 receives only noise, which may be discarded. The synchronous multiple access channel is likewise a special case of our general channel model.

For any transmitter–receiver pair  $(t, r)$ , the transfer function of the associated channel is

$$H_{(t,r)}(\omega) = \sum_{i=0}^m h_{(t,r)}[i] e^{-ji\omega}.$$

The spectral density of node  $r$ 's receiver noise process is

$$N_{(r)}(\omega) = \sum_{i=-(m-1)}^{m-1} R_{(r)}[i] e^{-ji\omega}.$$

We assume an average power constraint  $P_{(t)}$  for each transmitter  $t$  so that for all  $n$ ,

$$\frac{1}{n} \sum_{k=1}^n \mathbb{E} [X_{(t)}^2[k]] \leq P_{(t)}$$

for each  $t \in \{1, \dots, L\}$ , where the expectation is taken with respect to the distribution on the randomly chosen codeword  $(X_{(t)}[1], \dots, X_{(t)}[n])$ . For a given  $m$  we call this channel the linear Gaussian multiterminal channel (LGMC) with finite memory  $m$  and average power constraints  $(P_{(1)}, \dots, P_{(L)})$ . We define the initial state at transmitter  $t$  as

$$\mathcal{X}_{(t,0)} = (x_{(t)}[-m+1], \dots, x_{(t)}[0]).$$

This initial state, which determines the ISI from the previous codeword transmission to the current codeword transmission, impacts the first  $m$  channel outputs at each receiver. We define the initial channel state at receiver  $r$  as

$$\mathcal{V}_{(r,0)} = (V_{(r)}[-m+1], \dots, V_{(r)}[0]).$$

This initial state affects the first  $m$  channel outputs at the  $r$ th receiver due to noise correlation.

For any integers  $a, b \geq 1$ , let  $\mathbf{x}^a = (x_{(1)}^a, \dots, x_{(L)}^a)$  and  $\mathbf{x}_a^b = (x_{(1),a}^b, \dots, x_{(L),a}^b)$  where  $x_{(t)}^a = (x_{(t)}[1], \dots, x_{(t)}[a])$  and  $x_{(t),a}^b = (x_{(t)}[a], \dots, x_{(t)}[b])$ . Thus  $\mathbf{x}^a$  and  $\mathbf{x}_a^b$  describe channel inputs for all nodes  $1, \dots, L$  at times  $1, \dots, a$  and  $a, \dots, b$ , respectively. Define  $\mathbf{y}^a$  and  $\mathbf{y}_a^b$  similarly. We now define two  $n$ -block memoryless channel models derived from the LGMC. A multiterminal channel is  $n$ -block memoryless if for any integer  $K$

$$p(\mathbf{y}^{Kn} | \mathbf{x}^{Kn}) = \prod_{k=1}^K p(\mathbf{y}_{(k-1)n+1}^{kn} | \mathbf{x}_{(k-1)n+1}^{kn}). \quad (37)$$

For any such channel, the channel outputs  $y_{(r)}[k]$  in any  $n$ -block transmission are conditionally independent of channel inputs  $x_{(t)}[k]$  and noise samples  $V_{(r)}[k]$  from the previous and subsequent  $n$ -blocks given the channel inputs of the current  $n$ -block. In addition, the outputs  $y_{(r)}[k]$  for  $1 \leq k \leq n$  are independent of all initial channel states  $\mathcal{X}_{(t,0)}$  and  $\mathcal{V}_{(r,0)}$  given the channel

inputs  $x_{(t)}[k]$  for  $1 \leq k \leq n$ . Note that if we send codewords of length  $l \neq Nn$  over an  $n$ -block memoryless multiterminal channel then the previous codeword transmission may affect the channel outputs corresponding to the current codeword transmission, since the  $n$ -block memoryless channel is not  $l$ -block memoryless if  $l$  is not an integer multiple of  $n$ .

The circular multiterminal channel model, defined for  $n > m$ , is an  $n$ -block memoryless channel obtained by modifying the LGMC with memory  $m$ . Specifically, the  $n$ -block circular Gaussian multiterminal channel ( $n$ -CGMC) over each  $n$ -block has real input  $\mathbf{x}^n = (x_{(1)}^n, \dots, x_{(L)}^n)$  which produces the real output  $\mathbf{y}^n = (\tilde{y}_{(1)}^n, \dots, \tilde{y}_{(L)}^n)$  where for any  $r \in \{1, \dots, L\}$ ,  $\tilde{y}_{(r)}^n = (\tilde{y}_{(r)}[1], \dots, \tilde{y}_{(r)}[n])$  with

$$\begin{aligned} \tilde{y}_{(r)}[k] &= \sum_{t=1}^L \sum_{i=0}^m h_{(t,r)}[i] x_{(t)}[(k-i)_n] + \tilde{V}_{(r)}[k] \\ &= \sum_{t=1}^L h_{(t,r)}[k] \otimes x_{(t)}[k] + \tilde{V}_{(r)}[k] \end{aligned} \quad (38)$$

where  $\otimes$  denotes circular convolution and  $((s))_n$  again equals  $s$  modulo  $n$  except when  $s$  is zero or an integer multiple of  $n$ , in which case  $((s))_n = n$ . The noise process over each  $n$ -block  $\{\tilde{V}_{(r)}[k]\}_{k=1}^n$  is defined as a stationary Gaussian process with  $E[\tilde{V}_{(r)}[k]] = 0$  and

$$\begin{aligned} E[\tilde{V}_{(r)}[k]\tilde{V}_{(r)}[l]] &\triangleq \tilde{R}_{(r)}[k-l] \\ &= R_{(r)}[k-l] + R_{(r)}[k-l+n] \\ &\quad + R_{(r)}[k-l-n]. \end{aligned}$$

Thus, the noise autocorrelation function is periodic for noise samples within an  $n$ -block. Noise samples from different  $n$ -blocks are independent since the channel is  $n$ -block memoryless. The  $n$ -CGMC inherits the LGMC's power constraints.

The  $n$ -block memoryless Gaussian multiterminal channel ( $n$ -MGMC), defined for  $n > m$ , is obtained from the LGMC with memory  $m$  by restricting its outputs. Specifically, the  $n$ -MGMC derived from the LGMC with memory  $m$  has real input vector  $\mathbf{x}^n$  over each  $n$ -block with corresponding output vector at node  $r$  given by

$$(\hat{y}_{(r)}[m+1], \dots, \hat{y}_{(r)}[n]) = (y_{(r)}[m+1], \dots, y_{(r)}[n]).$$

Thus, for any receiving node  $r$  and any  $n$ -block,  $\hat{y}_{(r)}[k]$  is undefined over the first  $m$  symbols of the  $n$ -block and equals  $y_{(r)}[k]$  elsewhere. Note that noise samples from different  $n$ -blocks of the  $n$ -MGMC are independent due to the fact that  $R_{(r)}[i] = 0$  for  $|i| \geq m$ . Like the  $n$ -CGMC, the  $n$ -MGMC inherits the LGMC's power constraints.

For any multiterminal channel, let  $\mathcal{X}$  denote the common input alphabet of all the transmitters and  $\mathcal{Y}$  denote the common output alphabet of all the receivers. For the LGMC,  $n$ -CGMC, and  $n$ -MGMC,  $\mathcal{X} = \mathcal{Y} = \mathcal{Z} = \mathbb{R}$ . For each  $t, r \in \{1, \dots, L\}$ , let  $M_{(t,r)}$  be a positive integer and let  $\mathcal{M}_{(t,r)} = \{1, \dots, M_{(t,r)}\}$ . Then an  $[\{M_{(t,r)}\}, l]$  code for

an  $L$ -user multiterminal channel with independent information contains a family of encoders

$$\begin{aligned} X_{(t)}^l: \mathcal{M}_{(t,1)} \times \mathcal{M}_{(t,2)} \times \dots \times \mathcal{M}_{(t,L)} &\rightarrow \mathcal{X}^l, \\ X_{(t)}^l(W_{(t,1)}, \dots, W_{(t,L)}) &= (X_{(t)}[1], \dots, X_{(t)}[l]) \end{aligned}$$

for transmitters  $t \in \{1, \dots, L\}$  and a family of decoders

$$\begin{aligned} d_{(r)}: \mathcal{Y}^l &\rightarrow \mathcal{M}_{(1,r)} \times \dots \times \mathcal{M}_{(L,r)} \\ d_{(r)}(\mathbf{Y}_{(r)}^l) &= (\hat{W}_{(1,r)}, \dots, \hat{W}_{(L,r)}). \end{aligned} \quad (39)$$

(For any  $t = r$ , encoder  $X_{(t)}^l(\cdot)$  and decoder  $d_{(r)}(\cdot)$  both sit at node  $t$ .) In our definitions for the encoders and decoders, we simplify our notation by assuming that node  $t \in \{1, \dots, L\}$  sends a message to itself; in practice, no such message would be sent. Using this code, the rate transmitted from node  $t$  to node  $r$  is

$$R_{(t,r)} = \frac{1}{l} \log M_{(t,r)}.$$

Setting  $M_{(t,t)} = 1$  gives  $R_{(t,t)} = 0$ .

Let

$$\mathbf{d}(\mathbf{Y}^l) = (\mathbf{d}_{(1)}(Y_{(1)}^l), \dots, \mathbf{d}_{(L)}(Y_{(L)}^l))$$

denote the full vector of all decoded messages, and similarly let

$$\mathbf{W} = ((W_{(1,1)}, \dots, W_{(L,1)}), \dots, (W_{(1,L)}, \dots, W_{(L,L)}))$$

denote the corresponding vector of transmitted messages. The average error probability of an  $[\{M_{(t,r)}\}, l]$  code for an  $L$ -user multiterminal channel with finite memory and initial channel states  $\{\mathcal{X}_{(t,0)}, \mathcal{V}_{(r,0)}\}$ , assuming a uniform message distribution on  $W_{(t,r)} \in \mathcal{M}_{(t,r)}$  for each  $(t, r)$ , is defined as

$$\begin{aligned} P_e^l(\{\mathcal{X}_{(t,0)}, \mathcal{V}_{(r,0)}\}) &= \left( \frac{1}{\prod_{t=1}^L \prod_{r=1}^L M_{(t,r)}} \right) \\ &\times \sum_{\mathbf{w}} \Pr[\mathbf{d}(\mathbf{Y}^l) \neq \mathbf{w} | \mathbf{W} = \mathbf{w}, \{\mathcal{X}_{(t,0)}, \mathcal{V}_{(r,0)}\}]. \end{aligned} \quad (40)$$

Note that for  $n$ -block memoryless channels, including the  $n$ -CGMC and the  $n$ -MGMC, an  $[\{M_{(t,r)}\}, Nn]$  code for any integer  $N$  has error probability that is independent of the initial channel states by (37).

A rate set  $\{R_{(t,r)}\}$  is achievable for an  $L$ -user multiterminal channel with finite memory if for every  $\gamma > 0$  there exists a sequence of  $[\{M_{(t,r)}(l)\}, l]$  codes for this channel with

$$\frac{1}{l} \log M_{(t,r)}(l) \geq R_{(t,r)} - \gamma$$

for all  $l$  and

$$\sup_{\{\mathcal{X}_{(t,0)}, \mathcal{V}_{(r,0)}\}} P_e^l(\{\mathcal{X}_{(t,0)}, \mathcal{V}_{(r,0)}\}) \rightarrow 0$$

as  $l \rightarrow \infty$ . The capacity region for a multiterminal channel is the closure of the set of achievable rates. Since we are using block codes, time sharing gives us convexity of these capacity regions without an explicit convex hull operation.

We denote the capacity region for the LGMC, obtained using the above definitions, by  $\mathcal{C}$ . We denote the capacity region for the  $n$ -CGMC by  $C_n$  and the capacity region of the  $n$ -MGMC by  $Q_n$ . In general, the  $n$ -CGMC and  $n$ -MGMC and their corresponding capacity regions  $C_n$  and  $Q_n$  vary with  $n$ .

The first lemma states that  $\mathcal{C}$  is contained in the capacity region of the  $n$ -CGMC in the limit of infinite  $n$ .

*Lemma 1:*

$$\mathcal{C} \subseteq \overline{\left( \liminf_{n \rightarrow \infty} C_n \right)}$$

where  $\overline{(\cdot)}$  denotes closure.

*Proof:* We must show that every rate set  $\{R_{(t,r)}\}$  achievable for the LGMC may be approached, to arbitrary accuracy, by a rate set achievable for all  $n$ -CGMC's with  $n$  sufficiently large. Consider any rate set  $\{R_{(t,r)}\} \in \mathcal{C}$ . Fix  $\epsilon > 0$  and  $\gamma > 0$ . Then we can find  $n_0 \geq m$  such that for all  $n \geq n_0$

$$\left( R_{(t,r)} - \frac{\gamma}{4} \right) \frac{n}{n+m} > R_{(t,r)} - \frac{\gamma}{2}. \quad (41)$$

Consider now an  $(n+m)$ -CGMC for any fixed  $n > n_0$ . For this channel we must find, for all  $l$  sufficiently large, an  $[\{M_{(t,r)}\}, l]$  code with error probability

$$\sup_{\{\mathcal{X}_{(t,0)}, \mathcal{V}_{(r,0)}\}} \tilde{P}_e^l(\{\mathcal{X}_{(t,0)}, \mathcal{V}_{(r,0)}\}) < \epsilon$$

and

$$\frac{\log M_{(t,r)}}{l} \geq R_{(t,r)} - \gamma \quad \forall t, r \in \{1, \dots, L\}.$$

Let us first assume that  $l = N(n+m)$ . Then, since  $\{R_{(t,r)}\} \in \mathcal{C}$ , for all  $N$  sufficiently large we can find an  $[\{M_{(t,r)}\}, Nn]$  code for the LGMC with

$$\sup_{\{\mathcal{X}_{(t,0)}, \mathcal{V}_{(r,0)}\}} P_e^{Nn}(\{\mathcal{X}_{(t,0)}, \mathcal{V}_{(r,0)}\}) < \epsilon$$

and

$$\frac{\log M_{(t,r)}}{l} \geq R_{(t,r)} - \frac{\gamma}{4} \quad \forall t, r \in \{1, \dots, L\}. \quad (42)$$

Let us denote this code by  $\mathcal{S}_{Nn}$ .

Now consider an  $[\{M_{(t,r)}\}, N(n+m)]$  code  $\mathcal{T}_{N(n+m)}$  for the  $(n+m)$ -CGMC derived from the  $\mathcal{S}_{Nn}$  code as follows. We design the  $\mathcal{T}_{N(n+m)}$  code to recreate the ISI of the LGMC. Specifically, the  $\mathcal{S}_{Nn}$  code can be broken down into  $N$  blocks, each containing  $n$  symbols (giving  $N$   $n$ -blocks). We break the  $\mathcal{T}_{N(n+m)}$  code into  $N$  blocks of  $(n+m)$  symbols each where the first  $n$  symbols in the  $i$ th block of the  $\mathcal{T}_{N(n+m)}$  code are the same as the first  $n$  symbols in the  $i$ th block of the  $\mathcal{S}_{Nn}$  code and the last  $m$  symbols in the  $i$ th block of the  $\mathcal{T}_{N(n+m)}$  code are the

last  $m$  symbols in the  $(i-1)$ th block of the  $\mathcal{S}_{Nn}$  code. The last  $m$  symbols in the first block of the  $\mathcal{T}_{N(n+m)}$  code are chosen arbitrarily (corresponding to an arbitrary initial state  $\mathcal{X}_{(t,0)}$  for the  $\mathcal{S}_{Nn}$  code). More precisely, if the encoders for code  $\mathcal{S}_{Nn}$  are given by

$$X_{(t)}^{Nn}(W_{(t,1)}, \dots, W_{(t,L)}) = (X_{(t)}[1], \dots, X_{(t)}[Nn])$$

then the corresponding encoders for code  $\mathcal{T}_{N(n+m)}$  are

$$\begin{aligned} \tilde{X}_{(t)}^{N(n+m)}(W_{(t,1)}, \dots, W_{(t,L)}) &= (X_{(t)}[1], \dots, X_{(t)}[n], X_{(t)}[-m+1], \dots, X_{(t)}[0], \\ &\quad X_{(t)}[n+1], \dots, X_{(t)}[2n], X_{(t)}[n-m+1], \dots, \\ &\quad X_{(t)}[n], \dots, X_{(t)}[(N-1)n+1], \dots, X_{(t)}[Nn], \\ &\quad X_{(t)}[(N-1)n-m+1], \dots, X_{(t)}[(N-1)n]). \end{aligned} \quad (43)$$

The decoders in the  $\mathcal{T}_{N(n+m)}$  code operate on the first  $n$  of the  $(n+m)$  channel outputs in each  $(n+m)$ -block using the same decoder mappings as in the  $\mathcal{S}_{Nn}$  code. Specifically, if  $d_{(r)}$  denotes the decoder mappings for the  $\mathcal{S}_{Nn}$  code (as given by (39)) then the decoder mappings of the  $\mathcal{T}_{N(n+m)}$  code are given by

$$\begin{aligned} \tilde{d}_{(r)}(\tilde{Y}_{(r)}^{N(n+m)}) &= d_{(r)}(\tilde{Y}_{(r),1}^n, \tilde{Y}_{(r),n+m+1}^{2n+m}, \dots, \tilde{Y}_{(r),(N-1)(n+m)+1}^{Nn+(N-1)m}) \\ &= (\hat{W}_{(1,r)}, \dots, \hat{W}_{(L,r)}). \end{aligned} \quad (44)$$

By using our code structure to recreate the ISI of the LGMC on the CGMC we effectively convert the circular convolution of the  $(n+m)$ -CGMC into the linear convolution of the LGMC. Since the noise on the  $(n+m)$ -CGMC is circular, if the  $\mathcal{T}_{N(n+m)}$  code uses the same decoders as the  $\mathcal{S}_{Nn}$  code then its error probability on the circular channel equals the error probability of the  $\mathcal{S}_{Nn}$  code on the linear channel, as we now explicitly show.

We assume the same (arbitrary) initial states  $\{\mathcal{X}_{(t,0)}, \mathcal{V}_{(r,0)}\}$  on the  $(n+m)$ -CGMC as on the LGMC. Consider an arbitrary message set  $\{w_{(t,r)}\}$ . The  $\mathcal{S}_{Nn}$  code, which is designed without knowledge of the initial channel states  $\{\mathcal{X}_{(t,0)}, \mathcal{V}_{(r,0)}\}$ , gives channel inputs

$$X_{(t)}^n(w_{(t,1)}, \dots, w_{(t,L)}) = (X_{(t)}[1], \dots, X_{(t)}[Nn])$$

and channel outputs  $Y_{(r)}^{Nn}$ , while the  $\mathcal{T}_{N(n+m)}$  code has channel input given by (43) and corresponding outputs  $\tilde{Y}_{(r)}^{N(n+m)}$ . In particular, the first  $n$  outputs of the  $(n+m)$ -CGMC with the  $\mathcal{T}_{N(n+m)}$  code are given by

$$\begin{aligned} \tilde{Y}_{(r)}[k] &= \sum_{t=1}^L \sum_{i=0}^m h_{(t,r)}[i] \tilde{X}_{(t)}[((k-i))_{n+m}] + \tilde{V}_{(r)}[k] \\ &= \sum_{t=1}^L \sum_{i=0}^m h_{(t,r)}[i] X_{(t)}[k-i] + \tilde{V}_{(r)}[k]. \end{aligned}$$



Since the first  $n$  outputs of the LGMC with the  $\mathcal{S}_{Nn}$  are given by

$$Y_{(r)}[k] = \sum_{t=1}^L \sum_{i=0}^m h_{(t,r)}[i] X_{(t)}[k-i] + V_{(r)}[k]$$

we have that  $\tilde{Y}_{(r)}[k] = Y_{(r)}[k] - V_{(r)}[k] + \tilde{V}_{(r)}[k]$ . The first  $n$  outputs of each subsequent  $(n+m)$ -block of the  $(n+m)$ -CGMC satisfy a similar equation. In particular,

$$\begin{aligned} \tilde{Y}_{(r),(i-1)(n+m)+1}^{in+(i-1)m} \\ = Y_{(r),(i-1)n+1}^{in} - V_{(r),(i-1)n+1}^{in} + \tilde{V}_{(r),(i-1)(n+m)+1}^{in+(i-1)m}, \\ 1 \leq i \leq N. \end{aligned} \quad (45)$$

Let us now consider the error probability  $\tilde{P}_e^{N(n+m)}$  of the  $\mathcal{T}_{N(n+m)}$  code, which is independent of the initial channel states  $\{\mathcal{X}_{(t,0)}, \mathcal{V}_{(r,0)}\}$  by (37). Let  $\tilde{\mathcal{E}}_r$  denote the error event

$$\begin{aligned} \tilde{\mathcal{E}}_r = & \left[ d_{(r)} \left( \tilde{Y}_{(r),1}^n, \tilde{Y}_{(r),n+m+1}^{2n+m}, \dots, \tilde{Y}_{(r),(N-1)(n+m)+1}^{Nn+(N-1)m} \right) \right. \\ & \neq (w_{(1,r)}, \dots, w_{(L,r)}) \Big| \\ & \left. (W_{(1,r)}, \dots, W_{(L,r)}) = (w_{(1,r)}, \dots, w_{(L,r)}) \right]. \end{aligned}$$

Then the probability of decoding error using decoder (44) at receiver  $r$  given the transmitted messages  $(w_{(1,r)}, \dots, w_{(L,r)})$  is

$$\tilde{P}_e^{N(n+m)}(r) = \Pr[\tilde{\mathcal{E}}_r].$$

Similarly, let  $\mathcal{E}_r$  denote the event that the  $r$ th receiver of the LGMC makes an error with the  $\mathcal{S}_{Nn}$  code for the same message set. Note that this error event is not independent of the initial states. Thus

$$\begin{aligned} \mathcal{E}_r = & \left[ d_{(r)} \left( Y_{(r)}^{Nn} \right) \neq (w_{(1,r)}, \dots, w_{(L,r)}) \right. \\ & \left. (W_{(1,r)}, \dots, W_{(L,r)}) = (w_{(1,r)}, \dots, w_{(L,r)}), \right. \\ & \left. \{\mathcal{X}_{(t,0)}, \mathcal{V}_{(r,0)}\} \right]. \end{aligned}$$

Abbreviating  $\int_{\mathbb{R}} \dots \int_{\mathbb{R}} f(v^k) dv_1 \dots dv_k$  by  $\int f(v^k) dv^k$  for  $v_i \in \mathbb{R}$ , we have

$$\begin{aligned} \Pr[\tilde{\mathcal{E}}_r] &= \int \Pr \left[ \tilde{\mathcal{E}}_r \mid \tilde{V}_{(r)}^{N(n+m)} = v^{N(n+m)} \right] \\ &\quad \times p \left( \tilde{V}_{(r)}^{N(n+m)} = v^{N(n+m)} \right) dv^{N(n+m)} \\ &\stackrel{(a)}{=} \int \Pr \left[ \tilde{\mathcal{E}}_r \mid \tilde{V}_{(r)}^n = v^n, \dots, \right. \\ &\quad \left. \tilde{V}_{(r),(N-1)(n+m)+1}^{Nn+(N-1)m} = v_{(N-1)(n+m)+1}^{Nn+(N-1)m} \right] \\ &\quad \times p \left( \tilde{V}_{(r)}^{N(n+m)} = v^{N(n+m)} \right) dv^{N(n+m)} \\ &\stackrel{(b)}{=} \int \Pr \left[ \tilde{\mathcal{E}}_r \mid \tilde{V}_{(r)}^n = v^n, \dots, \right. \\ &\quad \left. \tilde{V}_{(r),(N-1)(n+m)+1}^{Nn+(N-1)m} = v_{(N-1)(n+m)+1}^{Nn+(N-1)m} \right] \\ &\quad \times \prod_{i=1}^N p \left( \tilde{V}_{(r),(i-1)(n+m)+1}^{in+(i-1)m} = v_{(i-1)(n+m)+1}^{in+(i-1)m} \right) \end{aligned}$$

$$\begin{aligned} &\times dv^{N(n+m)} \\ &\stackrel{(c)}{=} \int \Pr \left[ \tilde{\mathcal{E}}_r \mid \tilde{V}_{(r)}^n = v^n, \dots, \right. \\ &\quad \left. \tilde{V}_{(r),(N-1)(n+m)+1}^{Nn+(N-1)m} = v_{(N-1)(n+m)+1}^{Nn+(N-1)m} \right] \\ &\quad \times \prod_{i=1}^N p \left( \tilde{V}_{(r),(i-1)(n+m)+1}^{in+(i-1)m} = v_{(i-1)(n+m)+1}^{in+(i-1)m} \right) \\ &\quad \times \prod_{i=1}^N p \left( \tilde{V}_{(r),in+(i-1)m+1}^{in+(i-1)m} = v_{in+(i-1)m+1}^{in+(i-1)m} \right) \\ &\quad \times dv^{N(n+m)} \\ &\stackrel{(d)}{=} \int \dots \int \left( \int \dots \int \left[ \Pr \left[ \tilde{\mathcal{E}}_r \mid \tilde{V}_{(r)}^n = v^n, \dots, \right. \right. \right. \\ &\quad \left. \tilde{V}_{(r),(N-1)(n+m)+1}^{Nn+(N-1)m} = v_{(N-1)(n+m)+1}^{Nn+(N-1)m} \right] \\ &\quad \times \prod_{i=1}^N p \left( \tilde{V}_{(r),(i-1)(n+m)+1}^{in+(i-1)m} = v_{(i-1)(n+m)+1}^{in+(i-1)m} \right) \\ &\quad \left. \left. \tilde{V}_{(r),in+(i-1)m+1}^{in+(i-1)m} = v_{in+(i-1)m+1}^{in+(i-1)m} \right) \right] \\ &\quad dv^n \dots dv_{(N-1)(n+m)+1}^{Nn+(N-1)m} \\ &\quad \times \prod_{i=1}^N p \left( \tilde{V}_{(r),in+(i-1)m+1}^{in+(i-1)m} = v_{in+(i-1)m+1}^{in+(i-1)m} \right) \Big) \\ &\quad dv_{n+1}^{n+m} \dots dv_{Nn+(N-1)m+1}^{N(n+m)} \\ &\stackrel{(e)}{=} \int \left( \int \left[ \Pr \left[ \tilde{\mathcal{E}}_r \mid \tilde{V}_{(r)}^n = \hat{v}^n, \dots, \right. \right. \right. \\ &\quad \left. \tilde{V}_{(r),(N-1)(n+m)+1}^{Nn+(N-1)m} = \hat{v}_{(N-1)n+1}^{Nn} \right] \\ &\quad \times \prod_{i=1}^N p \left( \tilde{V}_{(r),(i-1)(n+m)+1}^{in+(i-1)m} = \hat{v}_{(i-1)n+1}^{in} \right) \\ &\quad \left. \left. \tilde{V}_{(r),in+(i-1)m+1}^{in+(i-1)m} = \bar{v}_{(i-1)m+1}^{im} \right) \right] d\hat{v}^{Nn} \\ &\quad \times \prod_{i=1}^N p \left( \tilde{V}_{(r),in+(i-1)m+1}^{in+(i-1)m} = \bar{v}_{(i-1)m+1}^{im} \right) \Big) d\bar{v}^{Nn} \\ &\stackrel{(f)}{=} \int \left( \int \left[ \Pr \left[ \tilde{\mathcal{E}}_r \mid \tilde{V}_{(r)}^n = \hat{v}^n, \dots, \right. \right. \right. \\ &\quad \left. \tilde{V}_{(r),(N-1)(n+m)+1}^{Nn+(N-1)m} = \hat{v}_{(N-1)n+1}^{Nn} \right] \\ &\quad \times \prod_{i=1}^N p \left( \tilde{V}_{(r),(i-1)n+1}^{in} = \hat{v}_{(i-1)n+1}^{in} \right) \\ &\quad \left. \left. \tilde{V}_{(r),(i-1)n-m+1}^{(i-1)n} = \bar{v}_{(i-1)m+1}^{im} \right) \right] d\hat{v}^{Nn} \\ &\quad \times \prod_{i=1}^N p \left( \tilde{V}_{(r),in+(i-1)m+1}^{in+(i-1)m} = \bar{v}_{(i-1)m+1}^{im} \right) \Big) d\bar{v}^{Nn} \end{aligned}$$

$$\begin{aligned}
&\stackrel{(g)}{=} \int \left( \int \left[ \Pr \left[ \mathcal{E}_r \mid V_{(r)}^n = \hat{v}^n, \dots, \right. \right. \right. \\
&\quad \left. \left. \left. V_{(r), (n-1)n+1}^{Nn} = \hat{v}_{(n-1)n+1}^{Nn} \right] \right. \right. \\
&\quad \left. \times \prod_{i=1}^N p \left( V_{(r), (i-1)n+1}^{in} = \hat{v}_{(i-1)n+1}^{in} \mid \right. \right. \\
&\quad \left. \left. V_{(r), (i-1)n-m+1}^{(i-1)n} = \bar{v}_{(i-1)m+1}^{im} \right) \right] d\hat{v}^{Nn} \\
&\quad \left. \times \prod_{i=1}^N p \left( \tilde{V}_{(r), in+(i-1)m+1}^{i(n+m)} = \bar{v}_{(i-1)m+1}^{im} \right) \right) d\bar{v}^{Nm} \\
&\stackrel{(h)}{=} \int \left( \Pr \left[ \mathcal{E}_r \mid \{\mathcal{X}_{(t,0)}, \mathcal{V}_{(r,0)}\} \right] \right. \\
&\quad \left. \times \prod_{i=1}^N p \left( \tilde{V}_{(r), in+(i-1)m+1}^{i(n+m)} = \bar{v}_{(i-1)m+1}^{im} \right) \right) d\bar{v}^{Nm} \\
&= \Pr \left[ \mathcal{E}_r \mid \{\mathcal{X}_{(t,0)}, \mathcal{V}_{(r,0)}\} \right] \quad (46)
\end{aligned}$$

where (a) follows from the fact that the decoder mapping (44) only operates on

$$\tilde{Y}_{(r), (i-1)(n+m)+1}^{in+(i-1)m}, \quad 1 \leq i \leq N$$

and, therefore, the decoder decision is only a function of

$$\tilde{V}_{(r), (i-1)(n+m)+1}^{in+(i-1)m}, \quad 1 \leq i \leq N;$$

(b) follows from the  $(n+m)$ -block memoryless property of the noise  $\tilde{V}_{(r)}$ ; (c) follows from conditioning; (d) results from splitting the integral over  $v^{N(n+m)} \in \mathbb{R}^{N(n+m)}$  into two pieces; (e) follows from renaming the dummy variables in the integrals; (f) follows from the fact that the noise autocorrelation is circular

$$\tilde{R}_{(r)}[k-l] = R_{(r)}[k-l] + R_{(r)}[k-l+(n+m)] + R_{(r)}[k-l-(n+m)]$$

over each  $(n+m)$ -block, so that

$$\begin{aligned}
&p(\tilde{V}_{(r), (i-1)(n+m)+1}^{in+(i-1)m} = \hat{v}_{(i-1)n+1}^{in} \mid \\
&\quad \tilde{V}_{(r), in+(i-1)m+1}^{i(n+m)} = \bar{v}_{(i-1)m+1}^{im}) \\
&= p(V_{(r), (i-1)n+1}^{in} = \hat{v}_{(i-1)n+1}^{in} \mid \\
&\quad V_{(r), (i-1)n-m+1}^{(i-1)n} = \bar{v}_{(i-1)m+1}^{im});
\end{aligned}$$

(g) follows from (45) and the decoder definition (44); and (h) follows from properties of conditioning.

Therefore, for an arbitrary set of messages and initial channel states, the error probability at the  $r$ th receiver of the  $(n+m)$ -CGMC using the  $\mathcal{T}_{N(n+m)}$  code equals the error probability at the  $r$ th receiver of the LGMC using the  $\mathcal{S}_{Nn}$  code. Since by assumption  $P_e^{Nn}$  as given by (40) is less than  $\epsilon$  for any set of initial channel states  $\{\mathcal{X}_{(t,0)}, \mathcal{V}_{(r,0)}\}$  by combining (46) with (40) we get that the  $\mathcal{T}_{N(n+m)}$  code has error probability  $\tilde{P}_e^{N(n+m)} < \epsilon$  on the  $(n+m)$ -CGMC. The corresponding data rates are

$$\frac{\log M_{(t,r)}}{N(n+m)} = \frac{\log M_{(t,r)}}{Nn} \frac{Nn}{N(n+m)} \geq R_{(t,r)} - \frac{\gamma}{2} \quad \forall t, r \in \{1, \dots, L\}$$

by (41) and (42).

Let us now consider arbitrary values of  $l$

$$l = N(n+m) + n', \quad n' \in \{0, \dots, n+m-1\}.$$

We form an  $[\{M_{(t,r)}\}, N(n+m) + n']$  code for the  $(n+m)$ -CGMC by appending  $n'$  arbitrary symbols to the codewords of the  $[\{M_{(t,r)}\}, N(n+m)]$  code described above. The decoder at each receiver discards the last  $n'$  outputs that it receives. For every initial channel state the error probability of this code is the same as that of the  $[\{M_{(t,r)}\}, N(n+m)]$  code since the decoders operate on the same received symbols. The code rate equals

$$\frac{\log M_{(t,r)}}{N(n+m) + n'} = \frac{\log M_{(t,r)}}{N(n+m)} \frac{N(n+m)}{N(n+m) + n'} \geq R_{(t,r)} - \gamma \quad \forall t, r \in \{1, \dots, L\} \quad (47)$$

for  $l$  sufficiently large. For systems with multiple codeword transmissions the  $l$ -block code is transmitted at integer multiples of  $n+m$  to maintain the same channel properties on each transmission. Equivalently, for  $1 \leq n' \leq n+m-1$ , the  $l$ -block code is padded with  $(n+m-1)-n'$  zeros, so that the next codeword transmission starts at an integer multiple of  $n+m$ . This causes a negligible rate penalty for  $N$  sufficiently large as can be seen by substituting  $n+m-1$  for  $n'$  in (47). Thus, for all  $l$  sufficiently large, we have exhibited a code for the  $(n+m)$ -CGMC with rate  $R_{(t,r)}$  and arbitrarily small error probability.  $\square$

The next lemma shows that for sufficiently large  $n$  the capacity region of the  $n$ -MGMC is contained in  $\mathcal{C}$ .

*Lemma 2:*

$$\overline{\left( \bigcup_{n>m} Q_n \right)} \subseteq \mathcal{C}.$$

*Proof:* We must show that for any  $\gamma > 0, \epsilon > 0, n > m$ , and  $\{R_{(t,r)}\} \in Q_n$ , if we take  $l$  sufficiently large then there exists an  $[\{M_{(t,r)}\}, l]$  code for the LGMC with

$$\sup_{\{\mathcal{X}_{(t,0)}, \mathcal{V}_{(r,0)}\}} P_e^l(\{\mathcal{X}_{(t,0)}, \mathcal{V}_{(r,0)}\}) < \epsilon$$

and

$$\frac{\log M_{(t,r)}}{l} \geq R_{(t,r)} - \gamma \quad \forall t, r \in \{1, \dots, L\}.$$

In this case, every point in  $\bigcup_{n>m} Q_n$  is an achievable rate point for the LGMC and the desired result follows, since  $\mathcal{C}$  is the closure of the set of all achievable rate points.

We first show this to be true for  $l$  an integer multiple of  $n$ . We then show it is true for all  $l$ . Fix  $n > m$  and consider any  $\{R_{(t,r)}\} \in Q_n$ . Then for any  $\gamma > 0$  and  $\epsilon > 0$  we can find an integer  $N_0$  sufficiently large such that for all  $N \geq N_0$  there exists an  $[\{M_{(t,r)}\}, Nn]$  code for the  $n$ -MGMC with

$$\frac{\log M_{(t,r)}}{Nn} \geq R_{(t,r)} - \frac{\gamma}{2} \quad \forall t, r \in \{1, \dots, L\}$$

and  $P_e < \epsilon$ . Note that  $P_e$  is independent of the initial channel state  $\{\mathcal{X}_{(t,0)}, \mathcal{V}_{(r,0)}\}$  by (37).

Suppose now that we use this  $[\{M_{(t,r)}\}, Nn]$  code on the LGMC. Then the rate is unchanged. Moreover, the error probability for any initial state  $\{\mathcal{X}_{(t,0)}, \mathcal{V}_{(r,0)}\}$  satisfies

$P_e(\{\mathcal{X}_{(t,0)}, \mathcal{V}_{(r,0)}\}) < \epsilon$  since if the decoder at receiver  $r$  ignores the outputs

$$\{y_{(r)}[1], \dots, y_{(r)}[m], y_{(r)}[n+1], \dots, y_{(r)}[n+m], \dots, \\ y_{(r)}[(N-1)n+1], \dots, y_{(r)}[(N-1)n+m]\}$$

then the resulting error probability is independent of  $\{\mathcal{X}_{(t,0)}, \mathcal{V}_{(r,0)}\}$  and the situation is equivalent to decoding the  $[\{M_{(t,r)}\}, Nn]$  code on the  $n$ -MGMC, for which  $P_e < \epsilon$ . Clearly, the error probability of the best decoder does not increase if these outputs are used in the decoding process. Thus, for  $l = Nn$ , we have exhibited a code for the LGMC with the desired rate and arbitrarily small error probability for all initial states  $\{\mathcal{X}_{(t,0)}, \mathcal{V}_{(r,0)}\}$ , assuming  $N$  (and, therefore,  $l$ ) sufficiently large.

We now extend this coding scheme for the LGMC to arbitrary values of  $l$ . Specifically, let  $l = Nn + n'$  for  $n' \in \{0, \dots, n-1\}$  and any integer  $N$  sufficiently large. We form an  $[\{M_{(t,r)}\}, Nn + n']$  code for the LGMC by appending  $n'$  arbitrary symbols to the codewords of the  $[\{M_{(t,r)}\}, Nn]$  code described above. The decoder at each receiver discards the last  $n'$  outputs that it receives. For all initial channel states the error probability of this code is the same as that of the  $[\{M_{(t,r)}\}, Nn]$  code, since the decoders of the  $Nn$  and  $Nn + n'$  codes operate on the same received symbols. The code rate equals

$$\frac{\log M_{(t,r)}}{Nn + n'} = \frac{\log M_{(t,r)}}{Nn} \frac{Nn}{Nn + n'} \geq \left(R_{(t,r)} - \frac{\gamma}{2}\right) \left(\frac{Nn}{Nn + n'}\right) \\ \forall t, r \in \{1, \dots, L\}.$$

Thus, for  $N$  sufficiently large we have

$$\frac{\log M_{(t,r)}}{Nn + n'} \geq R_{(t,r)} - \gamma.$$

Since  $\gamma$  is arbitrary, this gives our desired result.  $\square$

Our next lemma shows that for sufficiently large  $n$ , the capacity region of the  $n$ -CGMC is contained in that of the  $(n+m)$ -MGMC.

**Lemma 3:** There exists  $n_0$  sufficiently large such that for any  $n \geq n_0$ ,  $C_n \subseteq Q_{n+m}$ .

*Proof:* We must show that for  $n$  sufficiently large, every achievable rate point  $\{R_{(t,r)}\}$  for the  $n$ -CGMC is also achievable for the  $(n+m)$ -MGMC, i.e., for any  $\gamma > 0, \epsilon > 0, n > m$ , and  $\{R_{(t,r)}\} \in C_n$ , if we take  $l$  sufficiently large, then there exists an  $[\{M_{(t,r)}\}, l]$  code for the  $(n+m)$ -MGMC with error probability less than  $\epsilon$  and

$$\frac{\log M_{(t,r)}}{l} \geq R_{(t,r)} - \gamma \quad \forall t, r \in \{1, \dots, L\}.$$

We first consider codes where  $l$  is an integer multiple of  $n+m$  and then treat the more general case.

Assume  $n \geq 2m$  and let  $\{R_{(t,r)}\} \in C_n$ . Then for any  $\gamma > 0, \epsilon > 0$ , and  $N$  sufficiently large there exists an  $[\{M_{(t,r)}\}, Nn]$  code for the  $n$ -CGMC and corresponding decoders  $\hat{d}_{(r)}$  with rate

$$\frac{\log M_{(t,r)}}{Nn} \geq R_{(t,r)} - \frac{\gamma}{2}$$

and error probability  $P_e < \epsilon$ , where  $P_e$  is independent of the initial channel state by (37). Let us denote this code by  $\mathcal{U}_{Nn}$ . For code  $\mathcal{U}_{Nn}$ , let

$$\tilde{X}_{(t)}^{Nn}(W_{(t,1)}, \dots, W_{(t,L)}) = (\tilde{X}_{(t)}[1], \dots, \tilde{X}_{(t)}[Nn])$$

for all  $t \in \{1, \dots, L\}$  denote the codewords corresponding to messages  $\{(W_{(t,1)}, \dots, W_{(t,L)})\}_{t \in \{1, \dots, L\}}$ . This code can be broken down into  $N$  blocks of  $n$  input symbols ( $n$ -blocks) that are transmitted over the  $n$ -block memoryless  $n$ -CGMC.

Consider the following code  $\mathcal{V}_{N(n+m)}$  derived from  $\mathcal{U}_{Nn}$  for the  $(n+m)$ -MGMC. In the  $\mathcal{V}_{N(n+m)}$  code we use our code design to convert the linear convolution of the  $(n+m)$ -MGMC into the circular convolution of the  $n$ -CGMC. Specifically, the encoder  $X_{(t)}^{N(n+m)}$  of the  $\mathcal{V}_{N(n+m)}$  code appends  $m$  additional symbols  $\tilde{X}_{(t)}^{N(n+m)}[(i-1)n+k], k = 1, \dots, m$  to the  $i$ th  $n$ -block  $\tilde{X}_{(t)}^{in}, (i-1)n+1$  of the  $\mathcal{U}_{Nn}$  code for all  $i$  such that  $1 \leq i \leq N$ , giving

$$\hat{X}_{(t)}^{N(n+m)}(W_{(t,1)}, \dots, W_{(t,L)}) \\ = (\tilde{X}_{(t),1}^n, \tilde{X}_{(t),1}^m, \tilde{X}_{(t),n+1}^n, \tilde{X}_{(t),n+1}^m, \dots, \\ \tilde{X}_{(t),(N-1)n+1}^n, \tilde{X}_{(t),(N-1)n+1}^m).$$

The  $\mathcal{V}_{N(n+m)}$  code can be broken down into  $N$  nonoverlapping  $(n+m)$ -block inputs that are transmitted over the  $(n+m)$ -block memoryless  $(n+m)$ -MGMC. Let  $\hat{Y}_{(r)}^{Nn}$  denote the outputs at receivers  $r \in \{1, \dots, L\}$  of the  $n$ -CGMC with inputs  $\tilde{X}_{(t)}^{Nn}, t \in \{1, \dots, L\}$ . Let

$$(\hat{Y}_{(r),m+1}^{n+m}, \hat{Y}_{(r),n+2m+1}^{2(n+m)}, \dots, \hat{Y}_{(r),(N-1)n+Nm+1}^{N(n+m)})$$

denote the outputs at the receivers  $r \in \{1, \dots, L\}$  of the  $(n+m)$ -MGMC with inputs  $\hat{X}_{(t)}^{N(n+m)}, t \in \{1, \dots, L\}$ . The decoders for the  $\mathcal{V}_{N(n+m)}$  code are given by

$$d_{(r)}(\hat{Y}_{(r),m+1}^{n+m}, \hat{Y}_{(r),n+2m+1}^{2(n+m)}, \dots, \hat{Y}_{(r),(N-1)n+Nm+1}^{N(n+m)}) \\ = \tilde{d}_{(r)}(\hat{Y}_{(r),n+1}^{n+m}, \hat{Y}_{(r),m+1}^n, \hat{Y}_{(r),2n+m+1}^{2(n+m)}, \hat{Y}_{(r),n+2m+1}^{2n+m}, \\ \dots, \hat{Y}_{(r),Nn+(N-1)m+1}^{N(n+m)}, \hat{Y}_{(r),(N-1)n+Nm+1}^{Nn+(N-1)m})$$

for  $r \in \{1, \dots, L\}$  where, as stated above,  $\tilde{d}_{(r)}$  is the decoder for the  $n$ -CGMC.

We now use the fact that over each  $(n+m)$ -block, linear convolution becomes block-length- $n$  circular convolution with our  $\mathcal{V}_{N(n+m)}$  codeword design. Thus the code  $\mathcal{V}_{N(n+m)}$  has the same rate and error probability on the  $(n+m)$ -MGMC as the  $\mathcal{U}_{Nn}$  code has on the  $n$ -CGMC, as we now explicitly show.

Consider the first  $(n+m)$ -block transmission  $\hat{X}_{(t),1}^{n+m} = (\tilde{X}_{(t),1}^n, \tilde{X}_{(t),1}^m)$  over the  $(n+m)$ -MGMC. The output corresponding to this input is given by

$$\hat{Y}_{(r)}[k] = \sum_{t=1}^L h_{(t,r)}[k] * \hat{X}_{(t)}[k] + V_{(r)}[k] \\ = \sum_{t=1}^L h_{(t,r)}[k] \otimes \hat{X}_{(t)}[k] + V_{(r)}[k] \quad (48)$$

$$= \sum_{t=1}^L h_{(t,r)}[k] \otimes \tilde{X}_{(t)}[((k))_n] + V_{(r)}[k] \quad (49)$$

$$\stackrel{d}{=} \sum_{t=1}^L h_{(t,r)}[k] \otimes \tilde{X}_{(t)}[((k))_n] + \tilde{V}_{(r)}[((k))_n] \quad (50)$$

$$= \tilde{Y}_{(r)}[((k))_n] \quad (51)$$

for  $m+1 \leq k \leq n+m$ , where  $\stackrel{d}{=}$  denotes equality in distribution, (48) follows from the fact that with input  $\hat{X}_{(t),1}^{n+m}$ , linear convolution becomes block-length- $n$  circular convolution since the first  $m$  input symbols are repeated as the last  $m$  input symbols, (49) follows from the fact that  $\hat{X}_{(t)}[k] = \tilde{X}_{(t)}[((k))_n]$  for  $1 \leq k \leq n+m$ , (50) follows from the fact that

$$\begin{aligned} (V_{(r)}[m], \dots, V_{(r)}[n+m]) \\ \stackrel{d}{=} (\tilde{V}_{(r)}[((m))_n], \dots, \tilde{V}_{(r)}[((n+m))_n]) \end{aligned}$$

since

$$\mathbb{E}[V_{(r)}[k]] = \mathbb{E}[\tilde{V}_{(r)}[((k))_n]] = 0$$

and for  $m+1 \leq l, k \leq n+m$ , and  $n \geq 2m$ ,  $R_r[k-l] = \tilde{R}_r[((l))_n - ((l))_n]$ ,<sup>6</sup> and (51) follows from (38). Thus we have that  $(\hat{Y}_{(r),n+1}^{n+m}, \hat{Y}_{(r),m+1}^n) \stackrel{d}{=} (\tilde{Y}_{(r)}^n)$ . Using the same argument for each  $(n+m)$ -block transmission of input symbols

$$\hat{X}_{(t),(i-1)(n+m)+1}^{i(n+m)} = (\tilde{X}_{(t),(i-1)n+1}^{in}, \tilde{X}_{(t),(i-1)n+1}^{(i-1)n+m}), \quad 1 \leq i \leq N$$

yields

$$\left( \hat{Y}_{(r),in+(i-1)m+1}^{i(n+m)}, \hat{Y}_{(r),(i-1)n+im+1}^{in+(i-1)m} \right) \stackrel{d}{=} \left( \tilde{Y}_{(r),(i-1)n+1}^{in} \right), \quad \forall 1 \leq i \leq N. \quad (52)$$

Therefore, for any  $r \in \{1, \dots, L\}$

$$\begin{aligned} \Pr(\tilde{d}_{(r)}(\tilde{Y}_{(r)}^{Nn}) \neq (W_{(1,r)}, \dots, W_{(L,r)})) \\ = \Pr(\tilde{d}_{(r)}(\tilde{Y}_{(r),1}^n, \dots, \tilde{Y}_{(r),(N-1)n+1}^{Nn}) \\ \neq (W_{(1,r)}, \dots, W_{(L,r)})) \\ = \Pr(\tilde{d}_{(r)}(\hat{Y}_{(r),n+1}^{n+m}, \hat{Y}_{(r),m+1}^n, \dots, \\ \hat{Y}_{(r),Nn+(N-1)m+1}^{N(n+m)}, \hat{Y}_{(r),(N-1)n+Nm+1}^{Nn+(N-1)m}) \\ \neq (W_{(1,r)}, \dots, W_{(L,r)})) \quad (53) \end{aligned}$$

where (53) follows from (52) and the fact that the noise on the  $n$ -CGMC is  $n$ -block memoryless and the noise on the  $(n+m)$ -MGMC is  $(n+m)$ -block memoryless. Thus, (53) implies that the decoders  $d_{(r)}$  for the  $(n+m)$ -MGMC have the same

<sup>6</sup>Specifically,

$$\begin{aligned} \mathbb{E}[V_{(r)}[((k))_n]V_{(r)}[((l))_n]] &= \tilde{R}_r[((k))_n - ((l))_n] \\ &\stackrel{(a)}{=} \tilde{R}_r[k-l] \\ &\stackrel{(b)}{=} R_r[k-l] \end{aligned}$$

where (a) follows from the periodicity of  $\tilde{R}$  over an  $n$ -block and (b) follows from the fact that for  $n \geq 2m$ ,  $\tilde{R}_r[k-l] = R_r[k-l]$  for  $|k-l| < n$ .

error probability  $P_e < \epsilon$  as the decoders  $\tilde{d}_{(r)}$  in the  $n$ -CGMC, and this error probability is independent of the initial channel state by (37). We thus have, for all  $N$  sufficiently large, an  $[\{M_{(t,r)}\}, N(n+m)]$  code for the  $(n+m)$ -MGMC with arbitrarily small error probability and rate

$$\begin{aligned} \frac{\log M_{(t,r)}}{N(n+m)} &\geq \frac{\log M_{(t,r)}}{Nn} \frac{Nn}{N(n+m)} \\ &\geq \frac{n}{n+m} \left( R_{(t,r)} - \frac{\gamma}{2} \right) \quad \forall t, r \in \{1, \dots, L\}. \end{aligned}$$

We can thus find  $n$  sufficiently large such that

$$\frac{\log M_{(t,r)}}{N(n+m)} \geq R_{(t,r)} - \gamma \quad \forall t, r \in \{1, \dots, L\}.$$

Moreover, for  $N$  and  $n$  sufficiently large we can generate an  $[\{M_{(t,r)}\}, N(n+m) + n']$  code  $1 \leq n' < n+m$  for this channel by appending  $n'$  arbitrary symbols to the  $[\{M_{(t,r)}\}, N(n+m)]$  code described above. The decoder for this appended code ignores the last  $n'$  channel outputs. This code has the same error probability as the  $[\{M_{(t,r)}\}, N(n+m)]$  code since the decoder operates on the same received symbols. The  $n'$  appended symbols result in a negligible rate decrease for  $N$  sufficiently large. For systems with multiple codeword transmissions the code is transmitted at integer multiples of  $n+m$  to maintain the same channel properties on each transmission. Equivalently, the code is padded with  $(n+m-1)-n'$  additional zeros. This padding causes a negligible rate decrease for  $N$  sufficiently large. Thus for all  $l$  sufficiently large we have exhibited an  $[\{M_{(t,r)}\}, l]$  code for the  $(n+m)$ -MGMC with rate  $\{R_{(t,r)}\}$  and arbitrarily small error probability. So  $\{R_{(t,r)}\}$  is an achievable rate point for the  $(n+m)$ -MGMC for  $n$  sufficiently large.  $\square$

These lemmas culminate in the following theorem, which equates the capacity regions for all three of our multiterminal channel models.

*Theorem 1:*

$$\overline{\left( \limsup_{n \rightarrow \infty} C_n \right)} = \overline{\left( \limsup_{n \rightarrow \infty} Q_n \right)} = \overline{\left( \bigcup_{n > m} Q_n \right)} \quad (54)$$

$$= \mathcal{C} = \overline{\left( \liminf_{n \rightarrow \infty} C_n \right)} = \overline{\left( \liminf_{n \rightarrow \infty} Q_n \right)}. \quad (55)$$

*Proof:* We have the following chain of inclusions:

$$\overline{\left( \limsup_{n \rightarrow \infty} C_n \right)} \subseteq \overline{\left( \limsup_{n \rightarrow \infty} Q_n \right)} \quad (56)$$

$$\subseteq \overline{\left( \bigcup_{n > m} Q_n \right)} \quad (57)$$

$$\subseteq \mathcal{C} \quad (58)$$

$$\subseteq \overline{\left( \liminf_{n \rightarrow \infty} C_n \right)} \quad (59)$$

$$\subseteq \overline{\left( \liminf_{n \rightarrow \infty} Q_n \right)}. \quad (60)$$

The first inclusion (56) follows from Lemma 3, (57) follows from the definition of  $\limsup$ , (58) follows from Lemma 2, (59)

follows from Lemma 1, and (60) follows from Lemma 3. Moreover, since  $\liminf R_n \subseteq \limsup R_n$  for any sequence of regions  $\{R_n\}$ , all inclusions (56)–(60) become equalities.

Note that the closure operations in (54) and (55) are necessary since, although  $C_n$  and  $Q_n$  are closed sets, their  $\liminf$  and  $\limsup$  are not necessarily closed. However, since  $C_n$  and  $Q_n$  are convex (by a time-sharing argument), it can be shown that their  $\liminf$  and  $\limsup$  are also convex, and thus we need not take an explicit convex hull operation on the limiting capacity regions in (54) and (55).

#### APPENDIX B

We now prove that for  $n \geq 2m$  the noise terms  $V'_k$  and  $W'_k$  are statistically independent real Gaussian random variables with mean zero and variances

$$\sigma_{V'_k}^2 = nN_v(2\pi k/n)|\tilde{H}_k|^{-2}$$

and

$$\sigma_{W'_k}^2 = nN_w(2\pi k/n)|\tilde{G}_k|^{-2}.$$

It will suffice to prove this for  $V'_k$  only, since the proof for  $W'_k$  is identical.

From (13) we have that  $\tilde{V}_k$  is a weighted sum of the  $\tilde{v}_1, \dots, \tilde{v}_n$ , which are zero-mean Gaussian random variables. Thus,  $\tilde{V}_k$  will be a zero-mean Gaussian random variable. It will also be periodic with period  $n$ , since it is an  $n$ -point DFT. Let us define  $C_k = 1/\tilde{H}_k$  and  $\hat{V}_k = C_k \tilde{V}_k$ . Since  $\tilde{H}_k$  is the  $n$ -point DFT of a real sequence,  $C_k$  is also periodic and  $C_k = C_{n-k}^*$  for  $1 \leq k < n$ . The sequence  $V'_k$ ,  $1 \leq k \leq n$  is obtained by applying the transform (16) to  $\hat{V}_k$ . Since  $\hat{V}_k$  is just a constant times  $\tilde{V}_k$ , the real and imaginary parts of  $\hat{V}_k$  are zero-mean real Gaussian random variables, and therefore so is  $V'_k$  by (16).

Since  $V'_k$  is a zero-mean Gaussian random variable for each  $1 \leq k \leq n$  we need only evaluate  $E[V'_k V'_i]$  for  $1 \leq k, i \leq n$  to fully characterize the distribution on  $(V'_1, \dots, V'_n)$ . By (16) this expectation is found by evaluating  $E[\hat{V}_k^R \hat{V}_i^R]$ ,  $E[\hat{V}_k^R \hat{V}_i^I]$ , or  $E[\hat{V}_k^I \hat{V}_i^I]$  for any  $k$  and  $i$  such that  $1 \leq k, i \leq n$ . Using the fact that  $V_k^R = 0.5(V_k + V_k^*)$  and  $V_k^I = -0.5j(V_k - V_k^*)$  we see that all of these expectations can be expressed in terms of  $E[\hat{V}_k \hat{V}_i]$  and  $E[\hat{V}_k \hat{V}_i^*]$  as [3, eqs. I.2a–I.2c]

$$\begin{aligned} E[\hat{V}_k^R \hat{V}_i^R] &= 0.25 \left( E[\hat{V}_k \hat{V}_i] + E[\hat{V}_k \hat{V}_i^*] + E[\hat{V}_k^* \hat{V}_i] + E[\hat{V}_k^* \hat{V}_i^*] \right) \end{aligned} \quad (61)$$

$$\begin{aligned} E[\hat{V}_k^R \hat{V}_i^I] &= -0.25j \left( E[\hat{V}_k \hat{V}_i] - E[\hat{V}_k \hat{V}_i^*] + E[\hat{V}_k^* \hat{V}_i] - E[\hat{V}_k^* \hat{V}_i^*] \right) \end{aligned} \quad (62)$$

$$\begin{aligned} E[\hat{V}_k^I \hat{V}_i^I] &= -0.25 \left( E[\hat{V}_k \hat{V}_i] - E[\hat{V}_k \hat{V}_i^*] - E[\hat{V}_k^* \hat{V}_i] + E[\hat{V}_k^* \hat{V}_i^*] \right). \end{aligned} \quad (63)$$

Moreover, since

$$E[\hat{V}_k^* \hat{V}_i^*] = (E[\hat{V}_k \hat{V}_i])^*$$

and

$$E[\hat{V}_k^* \hat{V}_i] = (E[\hat{V}_k \hat{V}_i^*])^*$$

we need only evaluate  $E[\hat{V}_k \hat{V}_i]$  and  $E[\hat{V}_k \hat{V}_i^*]$  to obtain (61)–(63). We have

$$\begin{aligned} E[\hat{V}_k \hat{V}_i] &= C_k C_i E[\tilde{V}_k \tilde{V}_i] \\ &= C_k C_i E \left[ \sum_{m=1}^n \sum_{l=1}^n \tilde{v}_m \tilde{v}_l e^{-j2\pi mk/n} e^{-j2\pi li/n} \right] \\ &= C_k C_i \sum_{m=1}^n \sum_{l=1}^n E[\tilde{v}_m \tilde{v}_l] e^{-j2\pi mk/n} e^{-j2\pi li/n} \\ &= C_k C_i \sum_{m=1}^n \sum_{l=1}^n \tilde{R}_{(r)}[l-m] e^{-j2\pi(mk+li)/n} \\ &= C_k C_i \sum_{m=1}^n \sum_{l=m+1}^{n-m} \tilde{R}_{(r)}[l-m] \\ &\quad \times e^{-j2\pi(li-mi+mi+mk)/n} \\ &= C_k C_i \sum_{m=1}^n \sum_{l=m+1}^{n-m} \left( \tilde{R}_{(r)}[l-m] e^{-j2\pi i(l-m)/n} \right) \\ &\quad \times e^{-j2\pi m(i+k)/n} \\ &\stackrel{(a)}{=} C_k C_i \sum_{m=1}^n \sum_{x=1-m}^{n-m} \left( \tilde{R}_{(r)}[x] e^{-j2\pi ix/n} \right) \\ &\quad \times e^{-j2\pi m(i+k)/n} \\ &\stackrel{(b)}{=} C_k C_i \sum_{m=1}^n \sum_{x=-(m-1)}^{(m-1)} \left( \tilde{R}_{(r)}[x] e^{-j2\pi ix/n} \right) \\ &\quad \times e^{-j2\pi m(i+k)/n} \\ &\stackrel{(c)}{=} C_k C_i \sum_{m=1}^n \sum_{x=-(m-1)}^{(m-1)} \left( R_{(r)}[x] e^{-j2\pi ix/n} \right) \\ &\quad \times e^{-j2\pi m(i+k)/n} \\ &\stackrel{(d)}{=} C_k C_i \sum_{m=1}^n N_v \left( \frac{2\pi i}{n} \right) e^{-j2\pi m(i+k)/n} \\ &= C_k C_i N_v \left( \frac{2\pi i}{n} \right) \sum_{m=1}^n e^{-j2\pi m(i+k)/n} \\ &\stackrel{(e)}{=} n|C_k|^2 N_v \left( \frac{2\pi k}{n} \right) \delta_{((k+i))_n} \end{aligned} \quad (64)$$

where (a) follows from a change of variables, (b) follows from the fact that for  $n \geq 2m$ ,  $\tilde{R}_{(r)}[x] = 0$  for  $m \leq x \leq n-m$ , (c) follows from the fact that for  $n \geq 2m$  and  $|x| < m$ ,  $\tilde{R}_{(r)}[x] = R_{(r)}[x]$ , (d) follows from the fact that  $N_v$  is the DFT of  $R_v$ , and (e) follows from the orthogonality of the exponential functions and the fact that  $C_k = C_{n-k}^* = C_{-k}$ . A similar derivation yields that

$$E[\hat{V}_k \hat{V}_i^*] = n|C_k|^2 N_v \left( \frac{2\pi k}{n} \right) \delta_{((k-i))_n}. \quad (65)$$

Now combining (16) with (61)–(65) we get that

$$\begin{aligned} E[V'_k V'_i] &= nN_v(2\pi k/n)|\tilde{H}_k|^{-2} \delta_{((k-i))_n} \\ &= nN_v(2\pi k/n)|\tilde{H}_k|^{-2} \delta_{k-i} \end{aligned}$$

where the last equality follows from the fact that  $1 \leq k, i \leq n$ . Thus, for  $n \geq 2m$ ,  $V'_k$  and  $V'_i$  are uncorrelated for  $k \neq i$  and are, therefore, independent with the desired variance.

## APPENDIX C

In this appendix we give the formula for the capacity region of the broadcast channel with finite ISI when the transmitter sends both common and independent information. As in the case of parallel degraded broadcast channels [10], this capacity region is characterized implicitly as the intersection of six regions. The capacity region formula below is obtained by first using Corollary 1 to show that the desired region equals the capacity region of the circular broadcast channel with common information that is defined in Section II. The capacity region of this circular channel is derived by first decomposing the  $n$ -CGBC using the DFT, which leads to a set of parallel independent degraded broadcast channels with common information. The capacity region of this channel is given by [10, eq. 37] as the intersection of six regions. We then apply a similar derivation as in Section V to this implicit capacity region to obtain the limiting region of the  $n$ -CGBC with common information as  $n \rightarrow \infty$ . This derivation yields the following formula for the capacity region of the LGBC with finite ISI:

$$\mathcal{C} = \{(R_0, R_1, R_2):$$

$$R_0 \leq \frac{1}{2\pi} \int_{\Omega_1} \log \left( 1 + \frac{\bar{\alpha}(\omega)P(\omega)}{N_1(\omega) + \alpha(\omega)P(\omega)} \right) d\omega \\ + \frac{1}{2\pi} \int_{\Omega_2} \log \left( 1 + \frac{\bar{\alpha}(\omega)P(\omega)}{N_1(\omega) + \alpha(\omega)P(\omega)} \right) d\omega$$

$$R_0 \leq \frac{1}{2\pi} \int_{\Omega_1} \log \left( 1 + \frac{\bar{\alpha}(\omega)P(\omega)}{N_2(\omega) + \alpha(\omega)P(\omega)} \right) d\omega \\ + \frac{1}{2\pi} \int_{\Omega_2} \log \left( 1 + \frac{\bar{\alpha}(\omega)P(\omega)}{N_2(\omega) + \alpha(\omega)P(\omega)} \right) d\omega$$

$$R_0 + R_1 \leq \frac{1}{2\pi} \int_{\Omega_2} \log \left( 1 + \frac{\bar{\alpha}(\omega)P(\omega)}{N_1(\omega) + \alpha(\omega)P(\omega)} \right) d\omega \\ + \frac{1}{2\pi} \int_{\Omega_1} \log \left( 1 + \frac{P(\omega)}{N_1(\omega)} \right) d\omega$$

$$R_0 + R_2 \leq \frac{1}{2\pi} \int_{\Omega_1} \log \left( 1 + \frac{\bar{\alpha}(\omega)P(\omega)}{N_2(\omega) + \alpha(\omega)P(\omega)} \right) d\omega \\ + \frac{1}{2\pi} \int_{\Omega_2} \log \left( 1 + \frac{P(\omega)}{N_2(\omega)} \right) d\omega$$

$$R_0 + R_1 + R_2 \leq \frac{1}{2\pi} \int_{\Omega_1} \log \left( 1 + \frac{P(\omega)}{N_1(\omega)} \right) d\omega \\ + \frac{1}{2\pi} \int_{\Omega_2} \log \left( 1 + \frac{\bar{\alpha}(\omega)P(\omega)}{N_1(\omega) + \alpha(\omega)P(\omega)} \right) d\omega \\ + \frac{1}{2\pi} \int_{\Omega_2} \log \left( 1 + \frac{\alpha(\omega)P(\omega)}{N_2(\omega)} \right) d\omega$$

$$R_0 + R_1 + R_2 \leq \frac{1}{2\pi} \int_{\Omega_2} \log \left( 1 + \frac{P(\omega)}{N_2(\omega)} \right) d\omega \\ + \frac{1}{2\pi} \int_{\Omega_1} \log \left( 1 + \frac{\bar{\alpha}(\omega)P(\omega)}{N_2(\omega) + \alpha(\omega)P(\omega)} \right) d\omega \\ + \frac{1}{2\pi} \int_{\Omega_1} \log \left( 1 + \frac{\alpha(\omega)P(\omega)}{N_1(\omega)} \right) d\omega,$$

where  $0 \leq \alpha(\omega) \leq 1$  and  $P(\omega)$  satisfies (34).

## ACKNOWLEDGMENT

The authors gratefully acknowledge R. Gallager and R. Cheng for discussions on correlated noise, A. El Gamal for discussions on common information, and S. Shamai and J. Thomas for pointing out [8]. They are also grateful to the anonymous reviewers for their detailed suggestions and comments.

## REFERENCES

- [1] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
- [2] T. M. Cover, "Broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 2–14, Jan. 1972.
- [3] W. Hirt and J. L. Massey, "Capacity of the discrete-time Gaussian channel with intersymbol interference," *IEEE Trans. Inform. Theory*, vol. 34, pp. 380–388, May 1988.
- [4] S. Verdú, "Multiple-access channels with memory with and without frame synchronism," *IEEE Trans. Inform. Theory*, vol. 35, pp. 605–619, May 1989.
- [5] R. S. Cheng and S. Verdú, "Gaussian multiaccess channels with ISI: Capacity region and multiuser water-filling," *IEEE Trans. Inform. Theory*, vol. 39, pp. 773–785, May 1993.
- [6] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [7] J. Wolfowitz, *Coding Theorems of Information Theory*. Berlin, Germany: Springer-Verlag, 1964.
- [8] D. Hughes-Hartogs, "The capacity of a degraded spectral Gaussian broadcast channel," Ph.D. dissertation, Information Systems Lab., Center for Systems Research, Stanford Univ., July 1975.
- [9] G. S. Poltyrev, "Capacity for a sum of certain broadcast channels," *Probl. Pered. Inform.*, vol. 15, no. 2, pp. 40–44, 1979.
- [10] A. A. El Gamal, "Capacity of the product and sum of two unmatched broadcast channels," *Probl. Pered. Inform.*, vol. 16, no. 1, pp. 3–23, 1980.
- [11] A. V. Oppenheim and R. W. Schaffer, *Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1975.
- [12] P. Bergmans, "Random coding theorem for broadcast channels with degraded components," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 197–207, Mar. 1973.
- [13] W. Rudin, *Principles of Mathematical Analysis*. New York: McGraw-Hill, 1976.
- [14] D. G. Luenberger, *Optimization by Vector Space Methods*. New York: Wiley, 1969.